PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00, A01K 67/027, C12N 5/06, 5/10, C07K 14/47 | A1

(11) International Publication Number: WO 99/19469

(43) International Publication Date: 22 April 1999 (22.04.99)

(21) International Application Number: PCT/US98/21289

(22) International Filing Date: 9 October 1998 (09.10.98)

(30) Priority Data:
08/948,113 9 October 1997 (09.10.97) US

(71) Applicant: BIOTRANSPLANT, INC. [US/US]; Building 75, Third Avenue, Charlestown Navy Yard, Charlestown, MA 02129 (US).

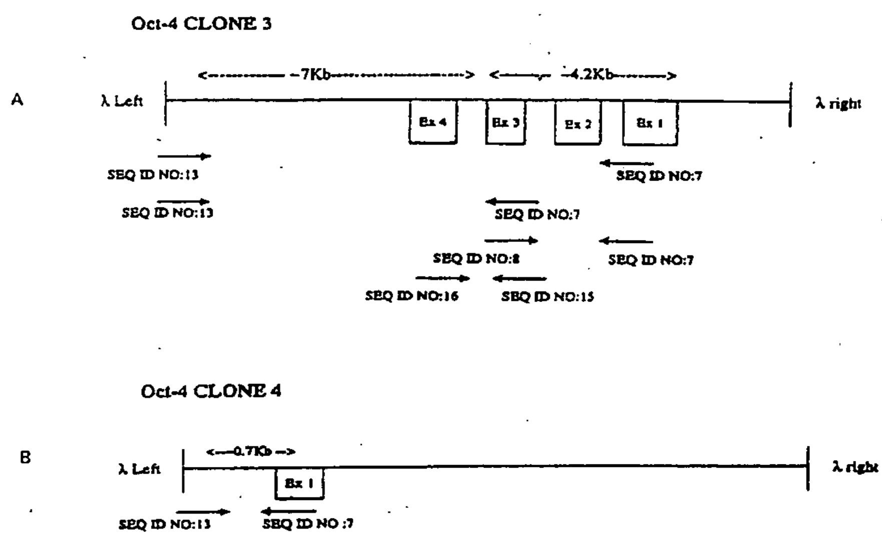
(72) Inventors: BAETSCHER, Manfred, W.; 6900 Southeast 35th Avenue, Portland, OR 97202 (US). AKIYOSHI, Donna, E.; 17 Hickory Lane, Upton, MA 01568 (US). KAPLAN, Ruth, A.; 10 Dunvegan Road, Tewsbury, MA 01876 (US).

(74) Agents: MULLINS, J., G. et al.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: PORCINE STEM CELLS COMPRISING A MARKER UNDER AN OCT-4 PROMOTER




(57) Abstract

The present invention provides for a method of isolating and/or propagating porcine stem cells, more specifically pluripotential porcine embryonic stem cells. The pluripotential cells are isolated and/or propagated by the use of a selectable marker gene which is inserted into the genetic material of the cells, and which permits the survival and growth of the porcine embryonic stem cells. The selectable marker gene is inserted so as to be regulated by a control or promoter polynucleotide sequence in the embryonic stem cells, for example the promoter polynucleotide sequence being the porcine Oct-4 promoter sequence of the present invention. The invention also provides for a transgenic pig which will constitute a source of the pluripotent cells.

AM

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# PORCINE STEM CELLS COMPRISING A MARKER UNDER AN OCT-4 PROMOTER

## Background Of The Invention

This invention relates to a method of isolating and/or enriching and/or selectively propagating pluripotential porcine cells, genetically modified porcine cells and pigs for use in said method, transgenic pigs providing a source of such cells and genetic selectable marker constructs for producing genetically modified cells and transgenic pigs.

Stem cells are progenitor cells which have the capacity both to self-renew and to differentiate into mature somatic cells. Embryonic stem cells are the archetypal stem cell, being capable of differentiating to form the whole gamut of cell types found in an adult animal. Such stem cells are described as "totipotential" or "pluripotential" as they are capable of differentiating into many cell types. Other types of stem cells, for example bone marrow stem cells and epidermal stem cells, persist in the adult animal. These stem cells have a more restricted capacity for differentiation.

In general, when required for research purposes or for medical use, stem cells have to be isolated from tissue samples by various fractionation procedures. However, even after careful segregation of cell types these stem cell preparations

consist of mixed cell types, and while enriched for stem cells include high proportions of differentiated cells which are not categorized as stem cells.

Furthermore, most stem cells cannot be grown readily in culture. When attempts are made to culture stem cells, the cells being cultured (which ordinarily contain a mixed population of cell types) grow at different rates and stem cells rapidly become overgrown by non-stem cell types. An exception is that embryonic stem cells from two specific strains of mice (129 and Black 6) can be cultured *in vitro* (Evans et al. (1981) Nature 292:154-156). Thus, established lines of murine embryonic stem cells can be obtained by culturing early (3 ½ day) embryonic cells from murine strain 129 and Black 6, or hybrids thereof. Embryonic cell lines from species other than the mouse are not so easily propagated. For an extensive review of isolation and propagation of stem cells see PCT publication WO 94/24274, incorporated by reference herein.

There has developed a pressing need to isolate and maintain in vitro embryonic stem cells from species of animals other than murine, such as other laboratory animals and domesticated animals, and most especially, from pigs. However, prior to the present invention the problems associated with producing cultures of porcine (or pig) stem cells, including the problem of producing cell populations of a satisfactorily low degree of heterogeneity and the problem of overgrowth in culture of non-pluripotent porcine cells, have not been solved. A particular problem associated with the continuing presence of certain differentiated cell types is that these can cause elimination of stem cells from the culture by inducing their differentiation or programmed cell death.

Thus according to the present invention, there is provided a porcine cell capable of being cultured under appropriate selective culture conditions so as to enable selective propagation of pluripotential stem cells, characterized in that said pluripotent porcine cells contain a genetic selectable marker, whereby a gene product associated with the genetic selectable marker is produced and which under said culture conditions causes selective survival and/or division of the desired pluripotent cells to occur. "Selective culture conditions" are those conditions under which a

population of cells is selectively grown. For example, to selectively grow cells that contain a gene which transfers resistance to a specific drug, the selective culture conditions would contain the drug so that all cells that did not express the drug resistance would be eliminated. Such selective culture conditions are well known in the art.

The invention further provides according to another aspect thereof, a transgenic animal, in this instance a transgenic pig, having genetic characteristics such that it or its progeny, during embryonic development or later life, constitute a source of porcine pluripotential cells as defined above. Such transgenic pigs may be produced according to the invention by introducing a genetic selectable marker into a fertilized oocyte or an embryonic cell, the genetic marker having the characteristics defined below, and utilizing the resulting transformed oocyte or embryonic cells as a progenitor cell for the desired transgenic animal.

A further aspect of the invention is vectors for use in producing an animal cell, for example a pig cell. Thus the invention further provides vectors for use in genetically modifying animal cells so as to produce transformed cells suitable for use as the source of cells for the method referred to below, said vector comprising a first genetic component corresponding to a genetic selectable marker and a second genetic component which, in the genetically modified porcine cell(s), results in the differential expression of the genetic selectable marker as a stably integrated construct. Such vectors may be in the form of expression vectors in which said second genetic component includes control sequences which are differentially activated in pluripotential stem cells and in cells other than the desired stem cells. The invention covers vectors which when used to transform porcine stem cells become integrated into the animal genome as well as vectors which do not become so integrated.

Summary Of The Invention

The present invention provides for a method of isolating and/or propagating porcine stem cells, more specifically pluripotential porcine embryonic stem cells. The pluripotential (or "pluripotent") cells are isolated and/or propagated by the use

of a selectable marker gene or nucleic acid sequence which is inserted into the genetic material of cells contained in a cell culture comprising porcine pluripotent embryonic stem cells, and which permits the survival and growth of said porcine embryonic stem cells. The selectable marker gene or nucleic acid sequence is inserted so as to be regulated by a control or promoter nucleotide sequence in said embryonic stem cells, for example the control sequence being the porcine Oct-4 promoter nucleotide sequence as described below. The terms "Oct-4 promoter nucleotide sequence" or "Oct-4 promoter sequence" refer to the promoter region of the Oct-4 gene, or any fragment of the promoter region that maintains promoter activity. The invention also provides for a transgenic pig which will constitute a source of said pluripotent cells.

By providing a sufficient and reliable source of porcine pluripotential embryonic stem cells, the present invention permits those skilled in the art to genetically modify the cells with a desired genetic modification. For example, said embryonic stem cells may be genetically altered so as to not express a cell surface membrane protein that may cause rejection of porcine cells after xenotransplantation. Said genetically altered cells are then useful in creating a transgenic pig, or line of transgenic pigs, which will not express said surface membrane protein and which, therefore, will contain organs that are less likely to be rejected upon xenotransplantation.

Brief Description Of The Figures

Figure 1 represents the genomic structure of murine Oct-4 (Yeom et al. (1991) Mechanisms of Development 35:171-179). Figure 1A shows a restriction map of an Oct-4 cosmid derived from the tw5g complete t-haplotype. The position of the transcription unit is indicated with a horizontal arrow running 5' to 3'. The position of the BamHI site missing in wild-type (C3H) and responsible for the restriction fragment length polymorphism (RFLP) between t and wild-type is marked with an asterisk. A vertical arrow indicates a BssHII site; NotI, NruI and MluI sites were absent. Figure 1B shows the exon/intron organization of the Oct-4 transcription unit.

Figures 2A, 2B, and 2C collectively show the polynucleotide sequence of F9 murine Oct-4 cDNA, and the deduced amino acid sequence, of murine Oct-4 protein. The numbering (right for the nucleic acid sequence; left for the amino acid sequence) begins at the putative initiation codon. The POU-specific domain and the POU-homeodomain are marked above the nucleotide sequence.

Figures 3A and 3B collectivelyshow the polynucleotide sequence of a portion of porcine Oct-4 exon 1 determined from genomic porcine DNA obtained from a commercial source (Clontech).

Figures 4A and 4B collectively show the polynucleotide sequence of a contiguous portion of porcine Oct-4 exon 1 determined from genomic porcine DNA from d/d haplotype miniswine.

Figure 5A illustrates the results obtained from PCR mapping of the d/d haplotype miniswine Oct-4 exon 1-containing Lambda clones #3.

Figure 5B illustrates the results obtained from PCR mapping of the d/d haplotype miniswine Oct-4exon 1-containingLambda clone #4..

Figures 6A and 6B collectively show the polynucleotide sequence of the porcine Oct-4-related sequence derived from Lambda clone #4.

Figure 7 shows the relationship between two fragments of porcine Oct-4, fragments 5D and 6A, derived from Lambda clone #3, and clone 12 (a PCR product encompassing a portion of exon 1 through exon 3). The letters A through H correspond to oligonucleotides used to map the fragments 5D and 6A derived from Lambda clone #3 and clone 12. Letter A corresponds to SEQ ID NO:7, letters B through H correspond to SEQ ID NO:17 through SEQ ID NO:23. The lengths of fragments derived for the various oligonucleotide pairs are as follows: G→F, 3.5Kb; A→G, 2.2Kb; F→H, 1.2Kb; B→F, 0.9Kb; D→E, 3Kb; H→E, 2.3Kb; and H→C, 3.5Kb.

Figures 8A, 8B, and 8C collectively show the promoter polynucleotide sequence for approximately 3 Kb 5' to the translation initiation codon for porcine Oct-4.

Figures 9A, 9B, and 9C collectively show an alignment comparison of the human, mouse and porcine Oct-4 promoter polynucleotide sequences. The sequence for the human Oct-4 promoter region includes nucleotides 1-499 from Genbank Accession Number Z11900; the sequence for the mouse Oct-4 promoter region includes nucleotides 1401-1950 from Genbank Accession Number S58422S1; the sequence for the porcine Oct-4 promoter region includes nucleotides 2701-3215 from SEQ ID NO:24. The major capping sites (RNA initiation sites) for the murine sequence (Okazawa et al. (1991) EMBO J. 10:2997-3005) are at nucleotides 480 and 501. The box marked as "A" is the SP1/HRE domain.

Figures 10A, 10B, and 10C collectively show an alignment comparison of the Retinoic Acid Responsive Element (RARE) regions of the pig and mouse Oct-4 promoter regions. The mouse sequence also shows high sequence identity within the region corresponding to S584221S1 nucleotides 430 through 1168 and porcine Oct-4 promoter sequence nucleotides 1534 through 2347. This region includes the retinoic acid responsive element located between nucleotides -1132 through -889 of Okazawa et al. (supra) (corresponding to nucleotides 609 through 1101 of Genbank Accession Number S584221S1). This region is also known as the "proximal enhancer region".

Detailed Description Of The Invention

There is an ever growing need for a sufficient and reliable source of pluripotential porcine embryonic stem cells. A particular desire for porcine embryonic stem cells is grounded in the need to be able to genetically modify the stem cells so that later cell development will result in an animal (pig) which will contain organs capable of being accepted immunologically by a recipient host. Once one is able to culture a large number of porcine stem cells and expand them in that state, then they can more easily be genetically modified to create a line of cells which will mature into organs less likely to be rejected after transplantation. These cells

can, according to the invention, be implanted to create a transgenic pig (and a pig line) which will contain organs useful for transplantation into a host.

Donor organ shortages have led to hopes that xenotransplantation could serve as an alternative means of organ availability. Swine, particularly miniswine, are an attractive alternative to nonhuman primate donors because of potentially greater availability, the reduced risk of zoonotic infections, appropriate size of organs and the reduced social and ethical concerns. However, one of the major barriers to xenotransplantation is hyperacute rejection. This phenomenon describes a very rapid and severe humoral rejection, which leads to destruction of the graft within minutes or hours of the transplant of the donor organ. Hyperacute rejection is apparently mediated by a complex series of events, including activation of the complement systems, activation of blood coagulation proteins, activation of endothelial cells and release of inflammatory proteins.

The hyperacute rejection process is initiated when the natural antibodies of the recipient bind to cells of the donor organ. The major cell surface protein (or epitope) that is recognized on porcine cells by human antibodies is the Gal$\alpha$1,3Gal $\beta$1,4GlcNAc structure. This structure is expressed at high levels on all mammalian cells, including swine cells, with the exception of human and old world non-human primates. A specific transferase, namely $\alpha$1,3 galactosyltransferase, is responsible for the transfer of a terminal galactose to the terminal galactose residue of N-acetyllactosamine-type carbohydrate chains and lactosaminoglycans. If one could eliminate or prevent the expression of the Gal$\alpha$1,3Gal $\beta$1,4GlcNAc epitope from the surface of the porcine cells in a transgenic pig, the cells would less likely be rejected by human antibodies and thus organs from such a transgenic pig would more likely be accepted upon transplantation.

In order to eliminate the $\alpha$1,3 galactosyltransferase activity in the pig, it would be highly desirable to have an abundant source of porcine embryonic stem cells, in order that sufficient genetic manipulations of the cells could be performed successfully. Using cultured porcine embryonic stem cells, a mutation, preferably a null mutation, is introduced by gene targeting at the native genomic locus encoding

$\alpha$1,3 galactosyltransferase. In order to permit such genetic manipulations in a desired quantity, porcine embryonic stem cells must be propagated in large amounts. Once the genetic manipulation is achieved (i.e., the $\alpha$1,3 galactosyltransferase activity is eliminated), the stem cells may be utilized in making a line of transgenic pigs. Such transgenic pigs would contain the genetically manipulated cells, which would not express the $\alpha$1,3 galactosyltransferase activity, and would therefore provide a supply of organs available for xenotransplantation with a minimized risk of rejection.

The isolation of embryonic stem cells from pigs would provide a method for the multiplication of animals with desired characteristics, an efficient means of producing transgenic animals, and a valuable model for studying cell development and differentiation. One of the major problems associated with the derivation of embryonic stem cells is that most stem cells cannot be grown readily in culture and when attempts are made to culture stem cells, the cells being cultured (usually a heterogeneous cell type population) grow at different rates and stem cells rapidly become overgrown by non-stem cell types. The present invention provides a method for isolating and/or enriching and/or selectively propagating pluripotential porcine embryonic cells which comprises maintaining a source of said cells under culture conditions conducive to cell survival, characterized in that the source of cells includes cells containing a genetic selectable marker which is is operatively linked to a porcine promoter nucleotide sequence which provides differential expression of the selectable marker in embryonic stem (pluripotential) cells and cells other than the desired stem cells, whereby differential expression of said genetic selectable marker results in preferential survival and/or division of the desired pluripotential porcine cells. Cell cultures and culture conditions suitable for propagating cells of the present invention are known by those skilled in the art. Suitable cell cultures may be, for example, found in Wurst, W. and A.L. Joyner, "Production of Targeted Embryonic Stem Cell Clones", in Gene Targeting: A Practical Approach (Ed. A.L. Joyner) 1993, Oxford Univ. Press; and Hashimoto et al., WO 95/34636.

In carrying out the method of the invention, the source of cells may include pluripotential cells containing a positive selectable marker and expression of the

marker is used to permit isolation and maintenance of the pluripotential cells. Alternatively (or additionally), the source of cells may include a negative selectable marker which is expressed in cells other than the desired pluripotential cells and is used to deplete the source of cells other than the desired pluripotential cells. The genetic selectable marker may, for example, be a foreign gene, a cellular gene or an antibiotic resistance gene such as for example the bacterial neomycin resistance gene.

It is preferred that the genetic selectable marker is operatively linked to a promoter nucleotide sequence which is differentially active in pluripotential stem cells and non-pluripotential cells. Promoter sequences may be included in the expression construct prior to introduction into the cells. The genetic selectable marker may be introduced into the source of cells by a variety of means known in the art including, but not limited to, injection, transfection, electroporation or by infection with a viral vector.

Further, the source of cells may be produced by transfection extemporaneously, or the source of cells may be derived from a transgenic animal, e.g., the founder transgenic animal or an animal at least one ancestor of which has had the aforementioned genetic marker introduced into its genetic complement. In such transgenic animals it is possible for the genetic selectable marker to pass down the germ line eventually resulting in the production of progeny, from the tissues of which the required source of cells can be derived.

A wide variety of known gene products may be relied upon for selective isolation and propagation of the desired porcine stem cells, including genetic selectable markers which are designed to protect the desired cells from the effects of an inhibiting factor present in the culture medium. In this instance, the inhibiting factor can, for example, be an antibiotic substance which inhibits growth or reproduction of cultured cells not expressing the gene (i.e., cells other than the desired pluripotential stem cells). Alternatively, the genetic selectable marker may selectively permit the growth of pluripotential stem cells using a marker known in the art. In this instance the marker may encode a growth factor, a growth factor

receptor, a transcription factor, or an immortalizing factor. Alternatively, the selectable marker may be a cell surface antigen or other gene product which allows purification or depletion of expressing cells for example by panning or fluorescence-activated cell sorting (FACS). The invention thus enables porcine stem cell populations to be obtained and/or maintained having a satisfactory degree of homology. Examples of all types of selectable markers discussed above are known in the art.

The present invention permits the development of expression constructs which direct specific expression of genetic selectable markers in porcine stem cells and not in differentiated cell types. Having introduced an expression construct by transfection or via the generation of transgenic animals, stem cells present within mixed cell populations can be isolated by culturing in the presence of the selection agent *in vitro*, or by otherwise manipulating the culture conditions by methods well known in the art.

One example of a gene which displays a suitably restricted stem cell expression pattern, and therefore should provide a suitable promoter polynucleotide sequence for use in controlling stem cell specific regulatory elements for the expression of a genetic selectable marker in accordance with the invention, is the Oct-4 gene. Octamer binding transcription factor 4 (hereinafter "Oct-4") is a member of the POU family of transcription factors. Oct-4 transcription, controlled by the Oct-4 promoter polynucleotide sequence, is activated between the 4 and 8-cell stage in the developing embryo and it is highly expressed in the expanding blastocyst and then in the pluripotent cells of the egg cylinder. Transcription is down-regulated as the primitive ectoderm differentiates to form mesoderm and by 8.5 days post coitum is restricted to migrating primordial germ cells. High level Oct-4 gene expression is also observed in pluripotent embryo carcinoma and embryonic stem cell lines and is down-regulated when these cells are induced to differentiate. As a result of the inventors' discovery of the sequence of the porcine Oct-4 promoter polynucleotide sequence, it is now possible to link the porcine Oct-4 promoter polyncleotide sequence to a genetic selectable marker in porcine cells and thus isolate and/or propagate a large number of porcine pluripotent stem cells.

As described herein, Applicants' Oct-4 promoter polynucleotide sequence is useful to control a genetic selectable marker sequence to permit propagation of large amounts of porcine embryonic stem cells. The Oct-4 promoter polynucleotide sequence may be used in its entirety, or a portion or fragment of the promoter sequence may be used in which the portion maintains the promoter activity. One skilled in the art would easily be capable of, using Applicants' sequence, joining a selectable marker to a portion or fragment of the porcine Oct-4 promoter polynucleotide sequence, transfecting the Oct-4-marker construct into a colony of cells, and growing the transfected cells in the appropriate medium to determine if the genetic selectable marker is translated. Within the scope of Applicants' invention are such portions or fragments that retain promoter activity, such portions having preferably at least 90% sequence identity, more preferably 95% sequence identity, and most preferably 98% sequence identity to the porcine Oct-4 promoter polynucleotide sequence as shown in Figure 8 or a portion of said Oct-4 promoter polynucleotide sequence, wherein said portion preferably consists of contiguous nucleotides from the porcine Oct-4 promoter polynucleotide sequence, i.e., a contiguous portion, and wherein said portion preferably comprises at least 38 nucleotides, more preferably at least 100 nucleotides, more preferably at least 200 nucleotides, more preferably at least 500 nucleotides, more preferably at least 1000 nucleotides, more preferably at least 1500 nucleotides, more preferably at least 2000 nucleotides, and most preferably at least 2500 nucleotides of the porcine Oct-4 promoter polynucleotide sequence shown in Figure 8.

Also within the scope of Applicants invention are polynucleotide sequences which hybridize to all or a portion of the Oct-4 promoter polynucleotide sequence as shown in Figure 8. Preferred under the scope of Applicants' invention are polynucleotide sequences which hybridize under high stringency conditions to all or a portion of the porcine Oct-4 promoter polynucleotide sequence as shown in Figure 8. Conditions under which hybridization will occur are known in the art and can be found in, for example, Bulletin 1234, Bio-Rad Laboratories, incorporated by reference herein.

The polynucleotides of the present invention may be in the form of DNA which DNA includes genomic DNA and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be either of the strands which together comprise the promoter. Such sequences are useful either in the promoter functionality or as probes to retrieve the promoter sequences.

Fragments of the full-length promoter of the present invention may be used as hybridization probes for the DNA containing the promoter or to isolate other DNAs which have a high polynucleode sequence identity to the promoter. Probes of this type preferably have at least 10 nucleotides, preferably at least 15 nucleotides, and even more preferably at least 30 nucleotides and may contain, for example, at least 50 or more nucleotides. In fact, probes of this type having at least up to 150 nucleotides or greater may be utilized. An example of a screen comprises isolating the promoter region of the porcine Oct-4 gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides, having a sequence complementary to that of the promoter or portion of the promoter sequences of the present invention are used to identify those polynucleotides that hybridize to, in a complementary sense, the promoter fragment, and have an identity as described above.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactive labels, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of the Oct-4 related promoter sequences from other sources or to screen such sources for related sequences.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences (SEQ ID NO:24 or a polynucleotide sequence encoding the same promoter as encoded by the sequence according to SEQ ID NO:24) if there is at least 70%, preferably at least 90%, and more preferably at leasst 95% identity between the sequences.

The polynucleotides of the present invention are preferably provided is an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment

Moreover, within the scope of Applicants' invention is the use of the porcine Oct-4 promoter polynucleotide sequence, or a portion thereof, connected to all or a portion of exon 1 and intron 1 of the porcine Oct-4 gene. The Oct-4 promoter plus exon/intron 1 is then joined to a genetic selectable marker according to the invention to be used in selectively propagating porcine stem cells.

Using applicants' teaching, pluripotential porcine embryonic stem cells are created in which the porcine Oct-4 promoter polynucleotide sequence is employed to drive stem cell specific transcription of a selectable marker. The stem cells are propagated in large quantities and are available for further genetic manipulation (e.g., to eliminate the $\alpha$1,3 galactosyltransferase activity). Applicants' invention also provides for transgenic pigs in which a porcine Oct-4 promoter polynucleotide sequence drives specific transcription of the selectable marker. An appropriate genetic selectable marker, for example, is the neomycin phosphotransferase gene which confers resistance to the antibiotic G418, although other genetic selectable markers are known and are available to those skilled in the art.

Selectable marker genes under the control of the porcine Oct-4 promoter may, according to the invention, be applied to the isolation of embryonic stem cell lineages from a transgenic pig.

Accordingly, this invention includes the generation of transgenic pigs which express the porcine Oct-4 promoter polynucleotide sequence linked to a genetic selectable marker, for example the selectable marker being $\beta$-geo (sequences encoding a fusion protein of $\beta$-galactosidase and neomycin phosphotransferase II). The $\beta$-galactosidase activity enables visualization of cells expressing the marker and the neomycin phosphotransferase activity confers resistance to the antibiotic G418.

Embryonic stem cells are then isolated from the transgenic pigs using culture conditions to support the growth of undifferentiated cells and maintaining selective pressure for cell expressing β-geo under the control of the Oct-4 promoter polynucleotide sequence.

Also included within the scope of the invention are transgenic pigs which contain cells expressing a genetic selectable marker under the control of the porcine Oct-4 promoter polynucleotide sequence according to the present invention, as well as transgenic pigs which contain cells expressing a genetic selectable marker under the control of the porcine Oct-4 promoter polynucleotide sequence which have been genetically manipulated to eliminate the α1,3 galactosyltransferase activity. The pigs containing the genetically manipulated cells are good resources for organs suitable for transplantation because of the minimal risk of hyperacute rejection.

In order to obtain the porcine Oct-4 promoter polynucleotide sequence, Applicants isolated porcine Oct-4 gene sequences and used those sequences to isolate and determine the DNA sequence upstream of the start of transcription initiation. The genomic map of the mouse Oct-4 gene (Yeom et al. (1991) Mechanisms of Development 35:171-179) is presented in Figures 1A and 1B. The entire sequence of the mouse Oct-4 gene (Schöler et al. (1990) Nature 344:435-439) is contained on a 4.73 Kb BamHI fragment (Figures 2A, 2B, and 2C). There are five exons and four introns. The sizes of the two short introns, 2 and 4, are 182 and 139 bp, respectively; the sizes of introns 1 and 3 are approximately 2430 and 450 bp, respectively. The POU domain is encoded by two or three split exons. The cap site of the murine Oct-4 transcript is located 163 bp upstream of the start codon (Yeom, 1991).

The invention will now be described in more detail in the following non-limiting examples with reference to the drawings. The examples are for illustration only and do not limit the scope of the present invention in any way, which is defined only by the claims. All references cited are hereby incorporated by reference herein in their entirety.

EXAMPLE 1 Cloning of Porcine Oct-4

1.1 Obtaining porcine Oct-4 exon 4 sequences

In order to obtain a fragment of porcine genomic DNA containing the Oct-4 sequence, degenerate PCR primers were designed based on sequence identity to several Oct family genes found in the Entrez database (http://www.ncbi.nlm.nih.gov/). These primers were designed such that they might amplify different Oct family members and were not specific for Oct-4. The sequences of the two degenerate primers are as follows:

DA34 (sense): 5' GCCCCTSCTGGAGAAGTGG 3' (SEQ ID NO:1)

DA37 (antisense): 5' GSCGSCGGTTRCAGAACCA 3' (SEQ ID NO:2)

DA34 maps within exon 3, which contains part of the POU-specific domain. DA37 maps within exon 4, which contains the POU-homeodomain.

Using a commercial source of porcine genomic DNA (Clontech, Palo Alto, CA, catalog # 6651-2) as template, a PCR fragment of approximately 1 Kb was amplified using Touchdown PCR (Roux, 1994) with Vent DNA Polymerase (New England Biolabs, Inc. ("NEB") Beverly, MA) on a Perkin-Elmer 9600 Thermocyler (Perkin-Elmer Corp., Norwalk, CT). Following NEB's protocol for setting up PCR reactions, the samples were amplified as follows: The tubes were heated at 94°C for 2 minutes followed by 3 cycles, each consisting of 45 seconds at 94°C, 1 minute at 55°C and 2 minutes at 72°C. The 3-cycle PCR step was repeated for 14 cycles during which the annealing temperature was sequentially decreased by 1°C every cycle. The final annealing temperature after 14 cycles was 40°C. Following this, 10 cycles were carried out at 94°C for 45 seconds, 1 minute at 40°C and 2 minutes at 72°C. A final extension for 5 minutes at 72°C completed the PCR program.

The 1 Kb PCR fragment was purified by gel electrophoresis followed by clean up on Qiaex resin (Qiagen, Chatsworth, CA) according to its manufacturer's protocol. The PCR fragment was treated with Taq polymerase in order to add the 3' A overhang. This allowed the cloning of this fragment into the TA vector, pCR II (Invitrogen, San Diego, CA). Several clones were sequenced using the primer

walking technique and standard DNA sequencing methodology. The complete sequence for clone pDA118-1 (960 bp) was determined. Based on sequence identity to both the human and mouse Oct-4, this clone was concluded to be authentic porcine Oct-4. The region of the Oct-4 gene contained within this clone extends from the last 48 bp of exon 3 to the first 27 bp of exon 5. Intron 3 was 628 bp, exon 4 was 159 bp and intron 4 was 99 bp in length.

1.2 Amplification, cloning and partial sequencing of a fragment from porcine genomic DNA which extends from exon 1 to exon 4.

A degenerate sense PCR primer within exon 1 was designed based on the sequence of human (Takeda et al. (1992) Nucleic Acids Research 20:4613-4620) and mouse Oct-4 (Okazawa et al. (1991) EMBO J. 10:2997-3005). The primer location was selected within the most conserved region of exon 1:

DA44 (sense): 5' CACCTGGCTTCRGAYTTCGCCTTC 3' (SEQ ID NO:3)

Using the DNA sequence derived from the 960 bp PCR fragment (clone PDA118-1), isolated from porcine genomic DNA (Clontech, Palo Alto, CA) and sequenced as described in 1.1 above, 2 antisense PCR primers were designed which map within either exon 3 or 4, respectively, and contain the following sequences:

DA43 (antisense, exon 3): 5' GCAGATTCTCGTTGTTGTCAGCTT 3' (SEQ ID NO:4)

DA58 (antisense, exon 4): 5' GTTGCCTCTCACTCGGTTCTCGATAC 3' (SEQ ID NO:5)

Using porcine genomic DNA (Clontech, Palo Alto, CA) as template, 2 PCR reactions were set up with either primer pair DA44/43 (SEQ ID NO:3/SEQ ID NO:4) or DA44/58 (SEQ ID NO:3/ SEQ ID NO:5) using the TaKaRa LA Taq Kit (Pan Vera, Madison, WI). PCR conditions were as recommended by the manufacturer, namely: 94°C for 2 minutes, 30 cycles at 98°C for 15 seconds, then 65°C for 11 minutes, followed by a 72°C extension for 5 minutes. The expected sizes of the PCR

products ranged from 3.2 to 6.8 kb. The observed PCR products were ~4.2 Kb for exon 1/3 primers and ~5.1 Kb for exon 1/4 primers.

The PCR products were purified by gel electrophoresis followed by clean-up on Qiaex resin (Qiagen, Chatsworth, CA). The purified fragments were cloned directly into the TA vector pCR II, transformed and bacteria colonies screened. Two clones were selected which contained either exon 1/3 fragment (pDA131-5) or exon 1/4 fragment (pDA132-9). Large-scale DNA preparations were made of each of these clones using Qiagen DNA Kit. Using standard DNA sequencing methodology that included the use of the SequiTherm Cycle Sequencing Kit (Epicentre Technologies, Madison, WI) and their vector sequencing primers, both clones were sequenced in from both ends. Sequence comparison of pDA132-9 with the sequence from pDA118-1 derived in Example 1.1 confirmed that this clone was authentic Oct-4. Exon 3 sequence of the two clones was also very similar. A consensus exon 1 sequence was obtained from sequence data of both clones. Approximately 340 bp of exon 1 sequence was determined for porcine Oct-4, as shown in Figures3A and 3B (SEQ ID NO:6). This sequence showed a high degree of sequence identity to both human and mouse Oct-4 sequences.

1.3 Obtaining an Oct-4 exon 1 probe

A second porcine Oct-4 probe was generated using the following oligonucleotide primers:

Oct-4 exon 1 5': 5' GGATCCTCGGACCTGGCTGAGCTTCCAA 3' (SEQ ID NO:7)

Oct-4 exon 3 3': 5' GAGCTCGTTGTTGTCAGCTTCCTCCACCCA 3' (inverse complement) (SEQ ID NO:8)

The oligonucleotide primer SEQ ID NO:7 includes nucleotides 92 through 118 of SEQ ID NO:6, while SEQ ID NO:8 was derived from the porcine exon 3 sequence.

Since the size in both the human and murine counterparts of porcine Oct-4, intron 1, is known to be quite large, i.e., 5 Kb and 2.43 Kb respectively (Takeda,

1992; Yeom, 1991), the PCRs were carried out using the TaKaRa LA Taq kit (Pan Vera Corp., Madison, WI). Each 50μl reaction contained 5 μl 10X LA PCR buffer II, 5 μl 25 mM $MgCl_2$, 8 μl dNTP mix (2.5 mM each), 2.5 units of TaKaRa LA Taq (0.5 μl), 500ng template DNA (porcine genomic DNA from Clontech (Palo Alto, CA) or from d/d miniswine #s 12021, 11378, 12023, 12037, 12038) and 100 pmol each oligonucleotide. The polymerase chain reaction (PCR) was performed using a Perkin Elmer DNA Thermal Cycler 480. The program used included 5 minutes at 94°C, followed by 14 cycles, each containing 20 seconds at 98°C and 8 minutes at 68°C. This was followed by 16 cycles, each consisting of 20 seconds at 98°C and 8 minutes at 68°C with an autoextension time of 15 seconds/cycle. The program concluded with a final extension of 5 minutes at 72°C and a 4°C soak. All reactions, except one, produced two PCR products: a major band at approximately 4.5 Kb and a minor band at 0.6 Kb (miniswine # 12037 gave only the smaller band). These were the sizes one would expect from an intron containing gene and an intronless gene, respectively, within this portion of Oct-4.

The 4.5 Kb PCR fragment obtained from miniswine #12038 was gel isolated using Qiaex II resin (Qiagen Inc., Chatsworth, CA) cloned into the vector pCR II and the resultant plasmid transformed into E. coli Inv ∝ F' using the instructions provided with the TA cloning kit (Invitrogen, San Diego CA). In order to identify those clones which contained the Oct-4 insert, two oligonucleotides were ordered from Genosys Biotechnologies (The Woodlands, TX) and colony PCRs were performed. The two oligonucleotides are shown below:

Oct-4 Ex 1- 5' seq: 5' CGGACCTGGCTGAGCTTCCAA 3' (SEQ ID NO:9)
Oct-4 Ex 1- 3' seq: 5' CCTCGGAGTTGCTCTCCACC 3' (SEQ ID NO:10)

SEQ ID NO:9 contains nucleotides 98 through 118 of SEQ ID NO:6, and SEQ ID NO:10 contains the reverse complement of nucleotides 307 through 326 of SEQ ID NO:6.

Bacteria derived from each of 40 white (presumed to be insert containing) colonies were separately inoculated into 50 μl of distilled water. The 40 samples were boiled for 5 minutes and centrifuged briefly in order to pellet the cells. Ten μl of each supernatant then served as template for PCRs which were carried out as

described previously for Taq DNA polymerase with the following exceptions: 1) the samples contained 50 μl rather than 100 μl and 2) the program used was 5' at 94°C, followed by 35 cycles, each containing 1 minute at 94°C, 1 minute at 55°C and 1 minute at 72°C. There was a final extension of 10 minutes at 72°C followed by a 4°C soak. The two oligonucleotides used in the PCRs should amplify a 230 bp exon 1 fragment of Oct-4. Most of the colonies tested were positive and one (#12) was chosen for sequence analysis. All sequencing was performed using either the Sequenase version 2.0 kit (United States Biochemical Corp., Cleveland, Ohio 44122) or the Fidelity DNA Sequencing kit (Oncor, Inc. Gaithersburg, MD 20877). Figures 4A and 4B (SEQ ID NO:11) show the sequence of the 5' end of clone #12, which includes most of exon 1 and a small portion of intron 1. By sequence comparison to the human and murine sequences, the exon/intron junction is estimated to be at nucleotide 320. The amino acid sequence of exon 1, which shows strong homology to both the murine and human genes, is underlined.

In order to prepare an exon 1 probe for use in a genomic Southern as well as in a subsequent library screen, the oligonucleotide shown above as SEQ ID NO:10 and the following oligonucleotide (SEQ ID NO:12) were used in a PCR for which the 4.5 Kb Oct-4 insert of clone #12 served as template:

Oct-4 ex 1 probe 5': 5' GGATCCTCGGACCTGGCTGAGC 3' (SEQ ID NO:12)

SEQ ID NO:12 includes nucleotides 92 through 112 of SEQ ID NO:6.

Ten identical PCRs, using Taq DNA polymerase, were carried out as described previously. A small portion of each reaction product was run on a 1% agarose gel in order to confirm the presence of a 236 bp band. The remainder of the reactions were then pooled, chloroform extracted, and ethanol precipitated. After resuspension in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), the DNA fragment was quantitated by running a portion on a 1% agarose gel along with markers containing known quantities of DNA.

1.4 Genomic Southern to verify utility of Oct-4 exon 1 probe

A genomic Southern was performed in order to determine whether or not the 236 bp probe, described in the previous section, would hybridize with specificity and thus could be safely used in a library screen (see Molecular Cloning, a Laboratory Manual, second edition, J. Sambrook et. al. , Cold Spring Harbor Laboratory Press, 1989). The genomic DNA used in both the Southern blot and library construction was obtained from d/d miniswine # 11852. For the Southern, 20 μg/reaction of genomic DNA was digested at 37°C overnight in a 50 μl volume with each of the following restriction enzymes: Bam HI, Bgl II, Eco RI, Hind III, Pst I, Sac I, Xba I (New England Biolabs, Inc., Beverly, MA). A small sample of each digest was run on a 0.7 % agarose gel in order to insure that the digests were complete. The remainder of each of the 7 reactions was then ethanol precipitated, rinsed in 70 % ethanol, dried, resuspended in 25 μl TE and run on a 15 X 15 cm 0.7 % agarose gel overnight at 20 V. A positive control was also included on 2 lanes of the gel, i.e., 2 pg and 20 pg respectively of the 4.5 Kb Oct-4 insert obtained by PCR, as previously described.

The next morning, the gel was soaked in 0.25 N HCl for 10 minutes, rinsed in water, and the DNA denatured for 45 minutes by soaking the gel in 1.5 M NaCl, 0.5 M NaOH. After a water rinse, the gel was placed for 45 minutes in a neutralization solution containing 1.5 M NaCl, 1M Tris-HCl pH 7.4 and then for another 45 minutes in 10X SSC (1.5 M sodium chloride/0.15 M sodium citrate, pH 7.0). The DNA was transferred onto a nitrocellulose membrane in 10X SSC overnight using the TurboBlotter Rapid Downward Transfer System (Schleicher & Schuell, Inc., Keene, NH 03431). After transfer, the nitrocellulose membrane was rinsed for 5 minutes in 5X SSPE (0.75 M sodium chloride, 0.05 M sodium phosphate, 0.005M EDTA) and then baked for 2 hours at 80°C in a vacuum oven.

Approximately 50 ng of the 236 bp probe was labeled by random priming with $\alpha$-$^{32}$P dCTP according to instructions provided with the High Prime DNA Labeling Kit (Boehringer Mannheim, Indianapolis, IN). The nitrocellulose blot was prehybridized at 65°C for 3 hours in a solution containing 6X SSC, 5X Denhardt's (0.5 g Ficoll, 0.5 g polyvinylpyrrolidone, 0.5 g bovine serum albumin), 100 μg/ml

sheared denatured herring sperm DNA, 0.75 % sodium dodecyl sulfate (SDS). Hybridization was carried out overnight at 65°C in the same solution to which the probe was added, following its denaturation in 1/10 volume 0.1 N NaOH. The blot was then washed 2 X 15 minutes at room temperature in 7X SSPE, 0.5 % SDS, 2 X 15 minutes at 37°C in 1X SSPE, 0.5 % SDS, 1 X 15 minutes at 65°C in 0.1X SSPE, 1 % SDS, rinsed in 0.1X SSPE at room temperature and exposed to X-ray film for 2 days at -80°C.

The results indicated that the Oct-4 exon 1 probe was quite specific since only 1 or 2 bands appeared for each of the restriction digests performed. Hybridization of the probe to the positive control was also seen, but only in the lane which contained the higher concentration (20 pg) of DNA.

1.5 Construction of a miniswine haplotype d/d genomic library

Miniswine genomic DNA (11852, haplotype d/d) was extracted from liver tissue and partially digested with Sau3AI. The Sau3AI sites were then partially filled-in and the fragments were size fractionated by gel electrophoresis using standard methodology. The genomic fragments were ligated to LambdaGEM-12 XhoI partially filled-in according to manufacturer's instructions (Promega, Madison, WI), packaged using Stratagene's Gold extract and titered on E. coli KW251 host cells.

The genomic library had 1.5 x $10^6$ independent clones, with an average insert size of ~12 Kb. The library was amplified one time, aliquoted and stored. The final titer of the amplified library was 1 x$10^{10}$ pfu/ml.

1.6 Screening the miniswine haplotype d/d genomic library for Oct-4 exon 1-containing clones

Approximately 3 X $10^5$ clones were screened with the Oct-4 exon 1 probe using duplicate nitrocellulose filters (Sambrook, 1989). The prehybridization, hybridization and wash conditions were essentially the same as those used for the Southern blot. The first round screen produced 5 positives; each of these was

replated at 3 different plaque densities and screened again with the Oct-4 exon 1 probe. Only 2 of the 5 were still positive after the second round screen. The two remaining clones (designated as #3 and #4 respectively) were plaque purified after a third round screen and λ DNA was prepared from plate lysates using a Qiagen Lambda Kit (Qiagen, Chatsworth, CA).

1.7 Mapping the Oct-4 exon 1-containing clones by PCR

Lambda clones #3 and #4 were mapped using LA PCR (Pan Vera, Madison, WI). The reactions were carried out as described previously except for a change in the program which increased all times at 68°C from 8 to 12 minutes. Two new oligonucleotides were made based on sequence provided by Promega (Madison, WI), the supplier of the LambdaGEM-12 vector:

RK EMBL3L: 5' GCAACGAACAGGTCACTATCAGTCA 3' (left arm of vector) (SEQ ID NO:13)

RK EMBL3R: 5' CTGCCTTCATTAAGGGCTGCGCAC 3' (right arm of vector) (SEQ ID NO:14)

It should be pointed out that the oligonucleotide shown as SEQ ID NO:14 never gave a positive PCR result, nor did two other oligonucleotides designed to amplify portions of the Oct-4 clones from the right arm of the vector. As can be seen in Figure 5, both clones are inserted in the vector with their 5' ends toward the right arm and their 3' ends toward the left arm of LambdaGEM-12. When the oligonucleotide pair of SEQ ID NO:7 and SEQ ID NO:13 was used, the size of the fragment amplified from clone #3 was approximately 11 Kb while the same pair amplified a much smaller 0.7 Kb fragment from clone #4. Additional PCRs using oligonucleotide pairs SEQ ID NO:15 and SEQ ID NO:13, SEQ ID NO:7 and SEQ ID NO:8, or SEQ ID NO:15 and SEQ ID NO:16 generated 7 Kb, 4.2 Kb, and 0.9 Kb products respectively from clone #3, but no products at all from clone #4. SEQ ID NO:15 includes sequences present in the porcine Oct-4 exon 4. SEQ ID NO:16 includes sequences in the exon 5/intron 4 junction region of porcine Oct-4.

Oct4 exon 3 5': 5'CTCGAGAAGTGGGTGGAGGAAGCGAC 3' (SEQ ID NO:15)

Oct4 exon 5/intron 4 3' : 5'GAATTCCACGCGGACCACCTGAGAG CAGGA3' (reverse complement) (SEQ ID NO:16)

These PCR results indicated that clone #3 contained the entire coding region of Oct-4 (i.e. exon 1 through exon 5 in addition to all 4 introns), several kilobases downstream, plus 2-3 kilobases upstream; clone #4 appeared to consist mostly of exon 1 and upstream regions.

1.8 DNA Sequence Analysis of Lambda Clones #3 and #4

In order to obtain exon 1 containing fragments of Lambda clones 3 and 4 for DNA sequence analysis, a Southern blot was prepared and probed with the exon 1 probe described earlier. The following restriction enzymes were used to digest 1-2 μg of clones 3 and 4 respectively, in 25 μl reactions: Bam HI, Eco RI, Hind III, Sac I, Xba I. All digests were carried out for 6 hours at 37°C and then loaded directly onto a 15 X 15 cm 1% agarose gel. The gel was run overnight at 20 V and the DNA transferred as described previously with the following changes: 1) instead of nitrocellulose, a positively charged nylon membrane was used, and 2) transfer was carried out in 20 X SSC instead of 10 X SSC.

The 236 bp exon 1 probe was labeled non-radioactively by random priming with DIG-11-dUTP. Labeling was carried out according to instructions provided with the Genius-2 DNA Labeling Kit (Boehringer Mannheim Corp., Indianapolis, IN). Following transfer, the DNA was affixed to the nylon blot by irradiation in a U.V. 1800 Stratalinker (Stratagene, La Jolla, CA). The blot was prehybridized for 3 hours at 50°C in DIG Easy Hyb (Boehringer Mannheim) and then hybridized overnight at 50°C in 15 ml of the same solution to which approximately 115 ng of boiled, denatured DIG-labeled probe had been added. Following hybridization, the blot was washed 2 X 15 minutes at room temperature in 2X SSC, 0.1% SDS, and then 2 X 15 minutes in 0.1X SSC, 0.1% SDS at 50°C. Oct-4 exon 1 containing bands were detected by chemiluminescence using the reagents Anti-Digoxigenin-alkaline phosphatase and the substrate CSPD, following instructions obtained from Boehringer Mannheim.

Four fragments which hybridized with the exon 1 probe, 2 from each Lambda clone (#3 and #4), were gel isolated using Qiaex resin and ligated into pBluescript KSII for sequencing. The resultant clones are described below:

1. 5D is a 5.5 Kb Bam HI fragment derived from Lambda clone #3. It contains exon 1, a portion of intron 1 and approximately 3 Kb upstream of exon 1.
2. 6A is a 4 Kb Sac I fragment derived from Lambda clone #3 which overlaps clone 5D. It contains exon 1, slightly more than 1 Kb upstream, and contains 1 Kb more of intron 1 downstream than does 5D.
3. 9-5 is a 1.3 Kb Hind III fragment derived from Lambda clone #4.
4. 10-C is a 0.6 Kb Sac I fragment of Lambda clone #4 which overlaps the Hind III clone above.

Sequence analysis of the overlapping fragments derived from Lambda clone #4 indicated that they were not derived from authentic Oct-4. The exon 1 sequence was related, but not identical to sequence obtained previously. There was no open reading frame in this exon 1-like DNA stretch. Furthermore it was joined to a sequence which was similar (but not identical) to that of exon 2, but with no intron 1 between them. A portion of this sequence can be seen in Figures 6A and 6B. Since this clone did not appear to contain authentic Oct-4, no further sequence analysis was carried out on $\lambda$ clone #4.

Figure 7 shows the relationship between fragments 5D and 6A derived from Lambda clone #3 and clone 12, described earlier. Mapping was done by sequence analysis as well as by LA PCR, using a number of oligonucleotides (SEQ ID NO:7 and SEQ ID NO:17 through SEQ ID NO:23) which are identified by the letters A through H on Figure 7. As can be seen, the 3 clones represent overlapping segments of porcine Oct-4 which extend from approximately 3 Kb upstream of exon 1 through exon 3. DNA sequence obtained from fragments 5D and 6A upstream of exon 1 further confirm that they are derived from authentic Oct-4. The murine counterpart of porcine Oct-4 (Oct3/4) lacks a TATA box, but possesses a putative Sp1/HRE (hormone-responsive element) in the proximal promoter region (Minucci et. al. (1996) EMBO J. 15:888-899). An identical sequence (see below) is found approximately 100 bp upstream of the translational start in the porcine gene, which

also lacks a TATA box: 5' GGGGGCGGGGCCAGAGGTCAAGGCTA 3'. The sequence determined for the promoter polynucleotide sequence of the porcine Oct-4 gene is shown in Figures 8A, 8B, and 8C (SEQ ID NO:24). This is what is referred to herein as Applicants' "Oct-4 promoter polynucleotide sequence" or the "Oct-4 promoter sequence". Figures 9A-C collectively show an alignment comparison of the human, mouse, and porcine Oct-4 promoter oligonucleotide sequences. Figures 10A-C collectively show an alignment comparison of the Retinoic Acid Responsive Element (RARA) regions of the porcine and murine Oct-4 promoter regions.

EXAMPLE 2 Testing of the Oct-4 Promoter

In order to demonstrate that the porcine Oct-4 promoter polynucleotide sequence is an embryonic stage specific promoter, the following constructs were made and tested in mouse embryonic stem cells.

2.1 Construction of pOct-4-βGeo

Plasmid pGT1.8 was obtained from Dr. P.S. Mountford (Stem Cell Sciences, Melbourne, Victoria, Australia). pGT1.8 contains a nuclear targeted Internal Ribosomal Entry Site (IRES) -βGeo construct. βGeo is a gene fusion of sequences encoding β-galactosidase and neomycin resistance gene (Friedrich et al. (1991) Genes & Development 5:1513-1523). 5' to the IRES is a sequence containing murine engrailed-2 splice acceptor which is designated as EN-2 (Gossler et al. (1989) Science 244:463-465). A 5.5 Kb fragment containing the porcine Oct-4 promoter, exon 1 and a portion of intron 1 was inserted into the SalI site of pGT1.8 as follows: Clone 5D (Example 1.8 above) was incubated with SalI/NotI and the 5.5 Kb DNA fragment was isolated. The DNA fragment was cleaved using Eco47III and the large 4.5 Kb 5'SalI/Eco47III DNA fragment was gel isolated.

A 0.6 Kb 3' portion of 5D was isolated after digestion with Eco47III/BamHI. The 3'BamHI site at the end of this fragment was converted to a XhoI site by PCR mutagenesis, using the following oligonucleotides:
PM Eco47III 5" 5' TCAAAGCGCTAAAATGTGATTTGG (SEQ ID NO:25)
PM XhoI3' 5' CGATCTCGAGGGATCCCAGACCGGGGAACT (SEQ ID NO:26)

The resulting PCR product was cloned into the TA vector, pCRII (Invitrogen, San Diego, CA) and sequenced. A correct clone was identified, cleaved with Eco47III and XhoI and gel purified. The final construct was constructed by ligation of dephosphorylated pGT1.8/SalI, the 4.5 Kb SalI/Eco47III 5' portion of 5D, and the 0.6 Kb Eco47III/XhoI 3' fragment. Since SalI and XhoI generate compatible overhanging ends, the Oct-4 insert could be inserted into pGT1.8 in either of two orientations. These orientations can be distinguished by XbaI cleavage. There are two XbaI sites in pGT1.8 and three in the Oct-4 insert: 125, 950, and 1990 bp from the SalI site. The correct orientation results in the following size fragments upon XbaI digestion: 5.3 Kb, 4.9 Kb, 3.0 Kb, 1.0 Kb, and 0.8 Kb. XbaI cleavage of the incorrect orientation generates 6.3 Kb, 4.9 Kb, 2.0 Kb, 1.0 Kb, and 0.8 Kb fragments.

2.2 Construction of pPGK-βGeo

The mouse PGK promoter was removed from a PGKneo vector (Aron Thall, BioTransplant, Inc.) using EcoRI and PstI. The resulting 0.5 Kb fragment was gel purified and the overhanging ends were filled in with Klenow DNA polymerase (NEB, Beverly, MA) according to instructions, phenol-chloroform extracted, and ethanol-precipitated. The DNA was resuspended in 1.5 μl $H_2O$, and 2 μg of phosphorylated HindIII linkers (2 μl) was added together with 1 μl 10 mM ATP, 1 μl 100 mM DTT, 2 μl 10X ligation buffer (Novagen, Madison, WI), and 1.5 μl T4 DNA ligase. The ligation reaction was incubated overnight at 16°C. The resulting DNA was digested with EcoR1 and HindIII and gel-purified. pOCUS-2 (Novagen, Madison, WI) was digested with EcoRI and HindIII and desphosphorylated using shrimp alkaline phosphatase (SAP). A ligation was performed with the pOCUS-2 fragment and the mouse PGK-promoter fragment. Following transformation of E. coli, colony PCRs were performed using primers that correspond to sequences located on either side of the multiple cloning site (MCS). The PCR template was DNA released from bacterial colonies which had been boiled for 5 minutes. A correct clone was grown up, plasmid DNA was prepared using a Qiagen column (Qiagen, Chatsworth, CA) and the DNA digested with HindIII and phosphatased.

A polynucleotide fragment containing the βGeo coding sequence was isolated from a vector provided by P.S. Mountford (Stem Cell Sciences, Melbourne, Australia). The plasmid was digested with ScaI and HindIII and the 4.3 Kb HindIII fragment containing βGeo and the polyadenylation site was gel isolated. A HindIII digest alone would have resulted in two fragments of approximately equal size, therefore ScaI was added in order to digest the unwanted DNA fragment into two smaller fragments.

A ligation was performed using the pOCUS-2/PGK promoter, linearized by HindIII and phosphatased, and the HindIII βGeo fragment. Following bacterial transformations, colony PCRs were carried out in order to identify a correct clone which contained the HindIII fragment in the correct orientation. The 5' PCR primer (5'AGCGCACGTCTGCCGCGCTGTT (SEQ ID NO:27)) is located in the PGK promoter, while the 3' primer (5'CCTGTAGCCAGCTTTCATCAAC (SEQ ID NO:28)) is in the βGeo fragment.

2.3 Construction of pPGK-EGFP

The EGFP derived plasmids were based upon the plasmid pEGFP-1 (Clontech, Palo Alto, CA; Genbank Accession #U55761) which encodes a variant of the green fluorescent protein used for monitoring the activity of promoters cloned into the multiple cloning site. EGFP is human codon-optimized and contains a chromophore mutation which produces fluorescence 35 times more intense than wild-type GFP. Sequences flanking the EGFP gene have been converted to a translation initiation consensus ribosome-binding site to further increase the translation efficiency in eukaryotic cells. The vector backbone provides an SV40 origin of replication and polyadenylation sequence, and a neomycin-resistance cassette for selection of stably transformed mammalian cells. One construct that served as the negative control was pEGFP without further modification.

pPGK-EGP was designed to contain the mouse PGK promoter driving expression of EGFP and was used as a positive control. The vector was constructed as follows: The vector pEGFP was digested with EcoRI and PstI and phosphatased. The PGK promoter was isolated by digestion of pPGKneo (Aron Thall,

BioTransplant, Inc.) using EcoRI and PstI. The two fragments, linearized pEGFP and the PGK promoter, were ligated. After transformation of E. coli, miniprep DNA was prepared (Qiagen, Chatsworth, CA) and correct clones were identified following digestion with EcoRI and PstI.

2.4 Construction of pOct-4-EGFP

The experimental construct, pOct-4-EGFP, contained the 3.2 Kb portion of the porcine Oct-4 promoter. The construct was made as follows: 5D was digested with BamHI, and the 5.5 Kb insert containing the porcine Oct-4 promoter sequence plus exon 1 and part of intron 1 was gel purified, and digested again with BstBI and FspI. The 2.84 Kb BamHI/BstBI fragment was gel purified. (The FspI was added to the digestion reaction in order to cleave an unwanted 3 kb fragment). The remaining 0.36 Kb portion of the promoter was obtained by PCR, using the 5D insert as template, and the following primers:

Oct4-BstBI 5' 5' CAGGGTCTTCGAAGAGGGGTCCA (SEQ ID NO:29)
Oct4-SalI 3' 5'GTCGACCAGGGCTCTCCAAGGGGA (SEQ ID NO:30)

This resulted in the addition of a SalI restriction site to the 3' end (SalI) which could be used for cloning purposes in pEGFP-1. Following PCR, the 0.36 Kb product was cloned into the TA vector (pCRII) and sequenced. After a correct clone was identified, a ligation was performed with the following DNA fragments: (1) pEGFP-1 digested with BamHI and SalI, and phosphatased; (2) the 2.84 Kb BamHI/BstBI portion of the Oct-4 promoter sequence; and (3) the 0.36 Kb BstBI/SalI 3' end of the Oct-4 promoter. Following transformation of E. coli, correct clones were identified by colony PCRs using the following primers:

5'GTCGACCAGGGCTCTCCAAGGGGA (SEQ ID NO:30)
6-T7 OL.4 5' 5'ACTTAGCACAGACACCASGACCT (SEQ ID NO:31)
A correct clone yields a 0.4 kb PCR product.

2.5 Cell culture

Mouse RW4 embryonic stem cells (ES) and mouse embryonic fibroblasts (MEFs) were obtained from Genome Systems (St. Louis, MO). The ES cells had been isolated from the inner cell mass of a 3.5 day embryo from the mouse strain 129/SvJ; the MEFs had been isolated from 14.5 day embryos. The ES cells were cultured in 80% DMEM, 15% FCS (#A-1115-L, Hyclone, Logan, UT), 2 mM L-glutamine, 0.1 mM non-essential amino acids, 10 mM HEPES, Pen/strep (100 units/ml), and 72 μM 2-mercaptoethanol. The ES cells were maintained in the presence of murine leukemia inhibitory factor (LIF, final concentration 1000 U/ml, Gibco/BRL, Gaithersburg, MD). The MEFs were grown in 87% DMEM, 10% FCS, 2 mM L-glutamine, and 0.1 mM non-essential amino acids.

2.6 Transfections and Selection

Transfections were performed by electroporation as follows: The cells were washed twice with phosphate-buffered saline (PBS), trypsinized using Trypsin/EDTA (0.05%/0.53 mM), and centrifuged. Colonies of 0.4–1.0 x $10^7$ cells were transfected with 25 μg linearized DNA (sequences from sections 2.1, 2.2, 2.3 and 2.4 as described above) in 800 μl electroporation buffer (20 mM HEPES, pH 7.05, 137 mM NaCl, 5 mM KCl, 6 mM D-glucose, 0.7 mM $Na_2HPO_4$) in a 0.4 cm cuvette using 300 v and 500 μF, at room temperature. The cells were seeded in 100 mm dishes (2.5 x $10^6$ cells/dish), containing a confluent monolayer of 30 Gy irradiated MEF. 24 hours later, the media was exchanged with selection medium containing 470 μg/ml G418. On day 8, G418 resistant colonies were picked and transferred to a 96-well plate that contained 50 μl trypsin solution. After 10 minutes, the cells were transferred to a 48 cell plate that contained medium plus 240 μg/ml G418. After three days, the medium was changed to 470 μg/ml G418 containing medium.

2.7 X-Gal (5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside) Staining Assay

The medium was removed from the culture dishes and the cells were washed briefly with PBS. Fixation solution (0.2% glutaraldehye, 0.1 M phosphate buffer, pH 7.3 [4 volumes 0.1 M disodium orthophosphate + 21 volumes 0.1 M sodium dihydrogen orthophosphate], 2 mM $MgCl_2$, 5 mM EGTA) was added to the cells for 5 minutes at 4°C. The cells were washed three times for 10 minutes each time with wash solution (0.1 M phosphate buffer, pH 7.3, 2 mM $MgCl_2$). The staining solution was added (1 mg/ml X-Gal [Sigma, St Louis MO] final concentration prepared by dissolving 25 mg X-Gal in 500 μl N,N dimethyl formamide, mixed with 25 ml of washing solution with 41 mg potassium ferricyanide and 52.5 mg potassium ferrocyanide) overnight at 37°C. The stained plates were stored in 0.2% glutaraldehyde at 4°C.

2.8 Retinoic Acid Assay

A stock solution of all trans-retinoic acid (Sigma, St. Louis, MO) was prepared by dissolving it in absolute ethanol to a final concentration of 1 mM.

The cells were seeded in 24-well plates containing a feeder layer of irradiated MEFs. After 3 days, the media was adjusted to 0.5 μM retinoic acid. The medium was changed every 2-3 days. After 5 days, the medium was removed and fresh medium without retinoic acid was added to the cells. Cells were split if necessary to avoid confluency. The colonies were then stained using the X-gal procedure.

2.9 X-Gal staining of cells transfected with the various vectors

The results (Table 1) demonstrate that the number of colonies derived from transfections using the PGK and Oct-4 promoter-containing plasmids was 50–100 fold higher than the number of colonies derived from the transfection using the promoterless plasmid. Therefore, the DNA sequence shown in Figure 8, isolated from porcine genomic DNA, contains the porcine Oct-4 promoter sequence. The single clone that resulted from the transfection using the promoterless plasmid was demonstrated to express the βGeo fusion protein. This clone proved to be useful as a

control for subsequent experiments as a non-developmentally regulated promoter. As used herein the staining legends indicate the following: XXX indicates a strong stain, XX indicates a moderate stain, and X indicates a weak stain.

**Table 1 Transfection Efficiency and X-Gal Staining of Neomycin Resistant**

| VECTOR | TOTAL # COLONIES | X-GAL STAINING |
|---|---|---|
| pβGeo - no promoter control | 1 (1 colony/5 x $10^6$ cells) | XXX |
| pPGKβGeo (pOCUS-2) | 9 (24 colonies/$10^6$ cells) | XXX |
| pOct-4-βGeo (pGTl.8) | 90 (12 colonies/$10^6$ cells) | XX |

Legend: X-weak stain XX-moderate stain XXX-strong stain

2.10 X-Gal staining after treatment of cell cultures with retinoic acid

The results of the X-Gal staining assay using cells that had been treated with retinoic acid are shown in Table 2.

Cells grown derived from the single colony that was obtained from the pβGeo plasmid transfection differentiated in the absence of LIF. These differentiated colonies expressed the β-Geo fusion protein, further establishing that the promoter driving the expression of this protein was not regulated by the presence of retinoic acid in the culture medium.

Cells from the colonies that were derived from the pPGKβGeo transfection also underwent differentiation in the presence of retinoic acid in the culture medium, with only undifferentiated cells from one colony (A5) surviving as a mixed population of differentiated and undifferentiated cells. This result was expected due to the understanding that the PGK promoter is differentially regulated by the presence of retinoic acid (Sutherland et al. (1995) Gene Expression 4:265-279). The differentiated cells did not stain with X-Gal, indicating that indeed the PGK promoter was not active in these differentiated cells. These cells functioned in this assay as the

positive control. The cells from the pOct-4βGeo derived colonies also differentiated in the presence of retinoic acid, with the resulting cell population being a mixture of a few undifferentiated cells together with a lot of differentiated cells. The undifferentiated cells stained with X-Gal but the differentiated cells did not. These results indicate that the porcine Oct-4 promoter is differentially regulated by the presence of retinoic acid in the culture medium. (Since retinoic acid promotes differentiation of the cells, it results in the turning off of the porcine Oct-4 promoter sequence because the sequence is only active in embryonic stem cells and not in differentiated cells.)

Table 2 X-Gal Staining of Colonies after Treatment of Cell Cultures with Retinoic

| Vector | LIF | G418 | Name of Colony | Undifferentiated colonies/well | Differentiated colonies/well | X-gal staining of undifferentiated colonies | X-gal staining of differentiated colonies |
|---|---|---|---|---|---|---|---|
| pβGeo | - | - | A2 | NO | XXX | na | XX |
| pPGKβGeo | - | - | A2 | NO | XXX | na | NO |
| | | | A3 | XX | XXX | X | NO |
| | | | A5 | X | XXX | X | X |
| pOct4βGeo | - | - | F4/A | X | XXX | X | NO |
| | | | E3/A | NO | XXX | na | NO |
| | | | E5/A | XX | XXX | X | NO |
| | | | E8/A | X | XXX | X | NO |
| | | | B1/A | X | XXX | X | NO |
| | | | E6/A | X | XXX | X | NO |
| | | | A1/B | XX | XXX | X | NO |
| | | | D2/B | XX | XXX | X | NO |
| | | | C3/B | NO | XXX | na | NO |

Legend: X-weak stain XX-moderate stain XXX-strong stain

An alternate method for inducing differentiation of ES cells is to culture them in the absence of LIF. Colonies were grown in the absence of a feeder layer but in the presence of G418 selection, for 12 days, and then subjected to the X-Gal staining protocol. The results are presented in Table 3.

The effects of the absence of LIF upon the ability of the pβGeo derived cells to undergo differentiation was less pronounced than the retinoic acid induced differentiation, but cells that did differentiate retained the ability to express the βGeo fusion protein, as expected.

Cells derived from the pPGKβGeo transfection did undergo differentiation, but did not lose the ability to express genes from the PGK promoter and therefore survived in the presence of G418. These results confirmed differences in the mechanism and/or extent of differentiation of ES cells induced by the absence of LIF but in the presence of retinoic acid or G418. The differentiated cells retained the ability to be stained by the X-Gal protocol.

Cells derived from the pOct-4βGeo transfections were similarly able to differentiate in the absence of LIF but did not express the βGeo protein. The cells appear to be differentially sensitive to the presence of G418, since some cells were still capable of proliferation despite the down-reulation of the pig Oct-4 promoter.

Table 3 X-Gal Staining of Cells Grown in the Absence of LIF, but in the Presence of G418 Selection

| Vector | LIF | G418 | Name of Colony | Undifferentiated colonies/well | Differentiated colonies/well | X-gal staining of undifferentiated colonies | X-gal staining of differentiated colonies |
|---|---|---|---|---|---|---|---|
| pβGeo | - | + | A2 | XXX | X | XXX | X |
| pPGKβGeo | - | + | A2 | X | XXX | X | X |
| | - | + | A3 | X | XX | XX | X |
| | | - | A5 | XX | XX | X | X |
| pOct4βGeo | - | + | F4/A | X | XX | X | NO |
| | - | + | E3/A | X | XX | X | NO |
| | - | + | E5/A | X | XX | X | NO |
| | - | + | E8/A | X | XX | X | NO |
| | - | + | B1/A | XXX | XX | XXX | NO |
| | - | + | E6/A | XX | XX | XX | NO |
| | - | + | A1/B | XX | XX | X | NO |
| | - | + | D2/B | XX | XX | XX | NO |
| | - | + | C3/B | X | X | X | NO |

Legend: X-weak stain XX-moderate stain XXX-strong stain

2.11 Transfection of porcine Oct-4/EGFP in mouse ES cells

Linearized plasmid DNAs were co-transfected with pPGKneo in a 1(PGK-Neo):10 (pOct-4/EGFP) ratio. Cells were selected for the ability to grow in 470 μg/ml G418. On day 7 fluorescent positive colonies were picked. No colonies grew from the promoterless control transfection; six colonies grew from the pPGK-EGFP

transfection, representing six colonies/million cells plated; eleven colonies grew from the pOct-4-EGFP transfection, representing 3 colonies/million cells plated. Colonies were subjected to the retinoic acid differentiation protocol, namely the colonies were grown in 0.5 μM retinoic acid for 5 days, then 7 days without retinoic acid (growing on MEFs). After this treatment all of the colonies lost the ability to fluorescesce, thus confirming that the pig Oct-4 promoter is developmentally regulated and hence is an ES cell-specific promoter.

As described above, Applicants' have provided the porcine Oct-4 promoter polynucleotide sequence which enables one to selectively propagate large numbers of porcine embryonic stem cells. Once this is accomplished, one skilled in the art is then able to create a transgenic pig line which will comprise the cells which will express the selectable marker under the control of the porcine Oct-4 promoter sequence. Technology to create such a transgenic animal is known in the art and can be found, for example, in Transgenic Animal Technology, A Laboratory Handbook ((1994) ed., Carl A. Pinkert, Academic Press, Inc., San Diego, CA).

Additionally, one skilled in the art may propagate large numbers of stem cells by utilizing the selectable marker under the control of the porcine Oct-4 promoter polynucleotide sequence as described above, and then further genetically modify these cells to accomplish a desired activity or to eliminate an activity (e.g., to eliminate α1,3 galactosyltransferase activity). (See, e.g., Tearle et al. (1996) Transplantation 61:13-19 entitled "The α1,3 galactosyltransferase Knockout Mouse: Implications For Xenotransplantation."). These cells that have been further genetically modified are then capable of being used to generate a transgenic pig line containing cells, and organs, with the desired genetic activity or eliminated activity.

Although the invention has been described with reference to its preferred embodiments, other embodiments, can achieve the same results. Variations and modifications to the present invention will be obvious to those skilled in the art and it is intended to cover in the appended claims all such modification and equivalents and follow in the true spirit and scope of this invention.

We claim:

1. A method of isolating or selectively propagating porcine stem cells, wherein said method comprises introducing into a source of cells containing porcine stem cells a genetic selectable marker construct which is operatively linked to a porcine promoter polynucleotide sequence which provides differential expression of the selectable marker in stem cells and cells other than the desired stem cells, and which under appropriate culture conditions enables the selective isolation and/or propagation of the desired stem cells.

2. The method of claim 1 wherein said porcine stem cells are embryonic stem cells.

3. The method of claim 1 wherein said porcine stem cells are pluripotential stem cells.

4. The method of claim 1 wherein the promoter polynucleotide sequence comprises a porcine Oct-4 promoter polynucleotide sequence.

5. The method of claim 4 wherein the porcine Oct-4 promoter polynucleotide sequence comprises the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

6. The method of claim 4 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 90% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

7. The method of claim 4 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 95% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

8. The method of claim 4 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 98% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

9. The method of claim 4 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence which hybridizes under high stringency conditions to all or a contiguous portion of the Oct-4 promoter polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

10. A method of selectively isolating and/or propagating porcine stem cells, said method comprising culturing a source of cells under selective culture conditions, characterized in that the source of cells includes porcine stem cells containing a genetic selectable marker whereby a gene product associated with the genetic selectable marker is produced and which under said culture conditions causes selective reproduction of the desired stem cells to occur, and wherein the genetic selectable marker is operatively linked to a porcine promoter polynucleotide sequence that regulates expression, which promoter polynucleotide sequence is differentially active in stem and non-stem cells.

11. The method of claim 10 wherein the promoter polynucleotide sequence comprises a porcine Oct-4 promoter polynucleotide sequence.

12. The method of claim 11 wherein the porcine Oct-4 promoter polynucleotide sequence comprises the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

13. The method of claim 11 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 90% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

14. The method of claim 11 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 95% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

15. The method of claim 11 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 98% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

16. The method of claim 11 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence which hybridizes under high stringency conditions to all or a contiguous portion of the Oct-4 promoter polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

17. An Oct-4 promoter polynucleotide sequence comprising the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

18. A polynucleotide sequence capable of hybridizing under conditions of high stringency to all or a portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

19. A porcine stem cell(s) capable of being cultured under appropriate selective culture conditions so as to enable isolation and/or propagation of porcine stem cells, characterized in that said porcine stem cell(s) contains a genetic selectable marker wherein differential expression of the selectable marker in the desired stem cell(s) and cells other than the desired stem cell(s) enables selective survival or growth of the desired stem cell(s) to occur, and wherein said genetic selectable marker is operatively linked to a porcine promoter polynucleotide sequence which provides differential expression of the selectable marker in stem cell(s) and cells other than the desired stem cell(s).

20. The porcine cell(s) of claim 19 wherein the porcine promoter polynucleotide sequence comprises a porcine Oct-4 promoter polynucleotide sequence.

21. The porcine cell(s) of claim 20 wherein the porcine Oct-4 promoter polynucleotide sequence comprises the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

22. The porcine cell(s) of claim 20 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 90% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

23. The porcine cell(s) of claim 20 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 95% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

24. The porcine cell(s) of claim 20 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 98% sequence identity with all or a contiguousportion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

25. The porcine cell(s) of claim 20 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence which hybridizes under high stringency conditions to all or a contiguous portion of the Oct-4 promoter polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

26. A transgenic pig which comprises a source of porcine cells suitable for the isolation and/or propagation of stem cells by a method according to any of claims 10 to 16.

27. A transgenic pig generated using a cell obtained by a method claimed in any of claims 1 to 9.

28. The transgenic pig according to claim 27 comprising cells which include said genetic selectable marker.

29. A vector for use in genetically modifying cells so as to be suitable for use as the source of cells defined in any of claims 1 to 16, said vector comprising a first genetic component corresponding to said genetic selectable marker and a second genetic component which in the genetically modified animal cells directly or indirectly results in the said differential expression of the selectable marker, and wherein said second genetic component comprises a porcine promoter polynucleotide sequence which is differentially activated in stem cells and cells other than stem cells.

30. The vector of claim 29 wherein said porcine promoter polynucleotide sequence comprises a porcine Oct-4 promoter polynucleotide sequence.

31. The vector of claim 30 wherein said porcine Oct-4 promoter polynucleotide sequence comprises the nucleotide polysequence as shown in Figure 8 (SEQ ID NO:24).

32. The vector of claim 30 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 90% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

33. The vector of claim 32 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 95% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

34. The vector of claim 32 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence having at least 98% sequence identity with all or a contiguous portion of the polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

35. The vector of claim 32 wherein the porcine Oct-4 promoter polynucleotide sequence comprises a polynucleotide sequence which hybridizes under high stringency conditions to all or a contiguous portion of the Oct-4 promoter polynucleotide sequence as shown in Figure 8 (SEQ ID NO:24).

36. The vector of claim 29 wherein the genetic selectable marker is an antibiotic marker.

37. The vector of claim 31 wherein the genetic selectable marker is an antibiotic marker.

38. A method of preparing a transgenic pig, said method comprising the steps of

providing a blastocyst;

providing porcine cells according to any of claims 19 to 25;

introducing the porcine cells into said blastocyst;

transferring the blastocyst to a recipient; and

allowing an embryo to develop to an animal to enable germline transmission of the selectable marker.

39. The porcine cell(s) of any of claims 19 to 25, said porcine cell(s) further comprising a genetic manipulation to eliminate $\alpha 1,3$ galactosyltransferase activity of said cell(s).

40. A transgenic pig comprising a porcine cell(s) according to any of claims 19 to 25.

41. A transgenic pig comprising a porcine cell(s) according to claim 39.

42. An organ suitable for transplantation into a recipient, said organ taken from a transgenic pig according to claim 41.

FIG. 1A

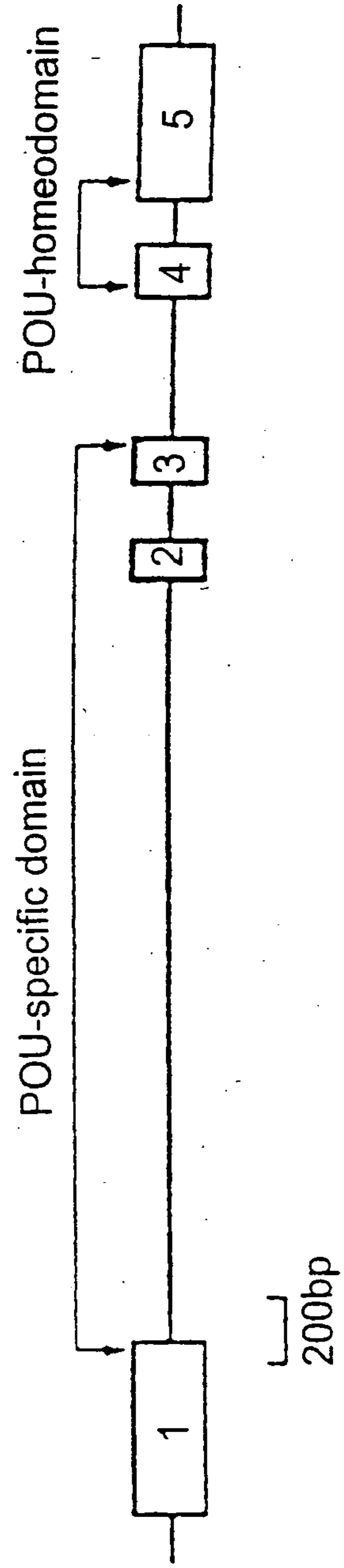



FIG. 1B

# FIG. 2A

```
                                                  BamHI SmaI  PstI EcoRI
                                                  AGTGGATCCCCCGGGCTGCAGGAATTCCGGG
     BamHI
     ATGGATCCTCGAACCTGGCTAAGCTTCCAAGGGCCTCCAGGTGGGCCTGGAATCGGACCA 60
1    M  D  P  R  T  W  L  S  F  Q  G  P  P  G  G  P  G  I  G  P  -

     GGCTCAGAGGTATTGGGGATCTCCCCATGTCCGCCCGCATACGAGTTCTGCGGAGGGATG 120
21   G  S  E  V  L  G  I  S  P  C  P  P  A  Y  E  F  C  G  G  M  -

     GCATACTGTGGACCTCAGGTTGGTCTGGGCCTAGTCCCCCAAGTTGGCGTGGAGACTTTG 180
41   A  Y  C  G  P  Q  V  G  L  G  L  V  P  Q  V  G  V  E  T  L  -
                                              Clone 16
     CAGCCTGAGGGCCAGGCAGGAGCACGAGTGGAAAGCAACTCAGAGGGAACCTCCTCTGAG 240
61   Q  P  E  G  Q  A  G  A  R  V  E  S  N  S  E  G  T  S  S  E  -

     CCCTGTGCCGACCGCCCCAATGCCGTGAAGTTGGAGAAGGTGGAACCAACTCCCGAGGAG 300
81   P  C  A  D  R  P  N  A  V  K  L  E  K  V  E  P  T  P  E  E  -
                        PstI  Clone 1
     TCCCAGGACATGAAAGCCCTGCAGAAGGAGCTAGAACAGTTTGCCAAGCTGCTGAAGCAG 360
101  S  Q  D  M  K  A  L  Q  K  E  L  E  Q  F  A  K  L  L  K  Q  -
```

FIG. 2B

```
                             POU-SPECIFIC DOMAIN
     AAGAGGATCACCTTGGGGTACACCCAGGCCGACGTGGGGCTCACCCTGGGCGTTCTCTTT 420
121  K  R  I  T  L  G  Y  T  Q  A  D  V  G  L  T  L  G  V  L  F  -

     GGAAAGGTGTTCAGCCAGACCACCATCTGTCGCTTCGAGGCCTTGCAGCTCAGCCTTAAG 480
141  G  K  V  F  S  Q  T  T  I  C  R  F  E  A  L  Q  L  S  L  K  -

     AACATGTGTAAGCTGCGGCCCCTGCTGGAGAAGTGGGTGGAGGAAGCCGACAACAATGAG 540
161  N  H  C  K  L  R  P  L  L  E  K  W  V  E  E  A  D  N  N  E  -

     AACCTTCAGGAGATATGCAAATCGGAGACCCTGGTGCAGGCCCGGAAGAGAAAGCGAACT 600
181  N  L  Q  E  I  C  K  S  E  T  L  V  Q  A  R  K  R  K  R  T  -

                                            POU HOMOEODOMAIN
     AGCATTGAGAACCGTGTGAGGTGGAGTCTGGAGACCATGTTTCTGAAGTGCCCGAAGCCC 660
201  S  I  E  N  R  V  R  W  S  L  E  T  M  F  L  K  C  P  K  P  -

     TCCCTACAGCAGATCACTCACATCGCCAATCAGCTTGGGCTAGAGAAGGATGTGGTTCGA 720
221  S  L  Q  Q  I  T  H  I  A  N  Q  L  G  L  E  K  D  V  V  R  -

     GTATGGTTCTGTAACCGGCGCCAGAAGGGCAAAAGATCAAGTATTGAGTATTCCCAACGA 780
241  V  W  F  C  N  R  R  Q  K  G  K  R  S  S  I  E  Y  S  Q  R  -
```

# FIG. 2C

GAAGAGTATGAGGCTACAGGGACACCTTTCCCAGGGGGGGCTGTATCCTTTCCTCTGCCC 840
261 E E Y E A T G T P F P G G A V S F P L P -

CCAGGTCCCCACTTTGGCACCCCAGGCTATGGAAGCCCCCACTTCACCACACTCTACTCA 900
281 P G P H F G T P G Y G S P H F T T L Y S -

GTCCCTTTTCCTGAGGGCGAGGCCTTTCCCTCTGTTCCCGTCACTGCTCTGGGCTCTCCC 960
301 V P F P E G E A F P S V P V T A L G S P -

ATGCATTCAAACTGAGGCACCAGCCCTCCCTGGGGATGCTGTGAGCCAAGGCAAGGGAGG 1020
321 H H S N * 324

TAGACAAGAGAACCTGGAGCTTTGGGGTTAAATTCTTTTACTGAGGAGGGATTAAAAGCA 1080
CAACAGGGGTGGGGGGTGGGATGGGGAAAGAAGCTCAGTGATGCTGTTGATCAGGAGCCT 1140
GGCCTGTCTGTCACTCATCATTTTGTTCTTAAATAAAGACTGGGACACACAGTAAAAAAA 1200
AAAAAAAAAAACTCGAG 1217
XhoI

FIG. 3A

```
        10         20         30         40         50
     *    *     *    *     *    *     *    *     *    *
GAATTCGGCT TCACCAGGCT TCGGACTTCG CCTTCTCGCC CCCGCCGGGC

        60         70         80         90        100
     *    *     *    *     *    *     *    *     *    *
GGTGGAGGCG ATGGGCCGGG AGGGCGGAGC CGGGCTGGGT TGATCCTCGG

       110        120        130        140        150
     *    *     *    *     *    *     *    *     *    *
ACCTGGCTGA GCTTCCAAGG GCCTCCCGGT GGGTCAGGGA TCGGGCCGGG

       160        170        180        190        200
     *    *     *    *     *    *     *    *     *    *
GGTTGGGCCG GGCGCCGAGG TGTGGGGGCT TCCCGCGTGC CCCCCGCCCT
```

# FIG. 3B

```
       210        220        230        240        250
     *    *     *    *     *    *     *    *     *    *
ATGACTTCTG CGGAGGGATG GCCTACTGCG CACCTCAGGT CGGAGTGGGG

       260        270        280        290        300
     *    *     *    *     *    *     *    *     *    *
CTGGTGCCCC AGGGCGGCCT GGAGACCCCT CAGCCCGAGG GCGAGGCGGG

       310        320        330        340
     *    *     *    *     *    *     *    *
GGCCGGGGTG GAGAGCAACT CCGAGGGGGC CTCCCCCGAG
```

FIG. 4A

```
                10          20          30          40          50
         *      *     *     *     *     *  *  *     *     *     *
GGATCCTCGG ACCTGGCTGA GCTTCCAAGG GCCTCCCGGT GGGTCAGGGA

                60          70          80          90         100
         *      *     *     *     *     *     *     *     *     *
TCGGGCCGGG GGTTGGGCCG GGCGCCGAGG TGTGGGGGCT TCCCGCGTGC

               110         120         130         140         150
         *      *     *     *     *     *     *     *     *     *
CCCCCGCCCT ATGACTTCTG CGGAGGGATG GCCTACTGCG CACCTCAGGT

               160         170         180         190         200
         *      *     *     *     *     *     *     *     *     *
CGGAGTGGGG CTGGTGCCCC AGGGCGGCCT GGAGACCCCT CAGCCCGAGG
```

# FIG. 4B

```
       210        220        230        240        250
    *    *     *    *     *    *     *    *     *    *
GCGAGGCGGG GGCCCGGGTG GAGAGCAACT CCGAGGGGGC CTCCCCCGAG

       260        270        280        290        300
    *    *     *    *     *    *     *    *     *    *
CCCTGTGCCG CCCCCGCTGG CGCCGCGAAG CTGGACAAGG AGAAGCTGGA

       310        320        330        340        350
    *    *     *    *     *    *     *    *     *    *
GCCGAACCCC GAAGAGGCGA GTGAGCTGCC GGGAGCTGGG GGAGGCGATC

       360        370        380        390        400
    *    *     *    *     *    *     *    *     *    *
GCGCTGGCCG GGGGCGCACG CAGGGGAGGT GGTCGCCTGC CGCCCGGGCA

    *
GGAGGGG
```

FIG. 5A

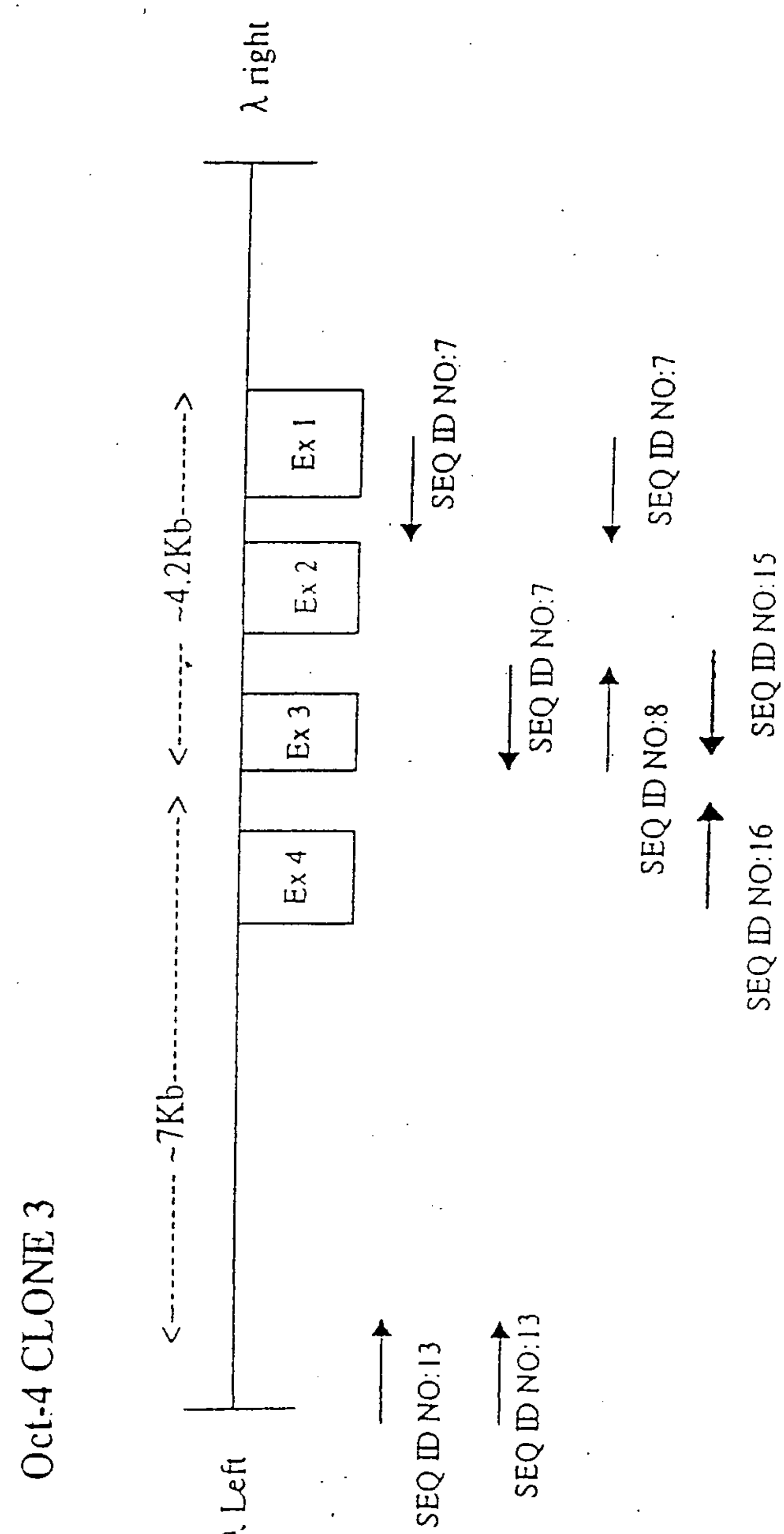
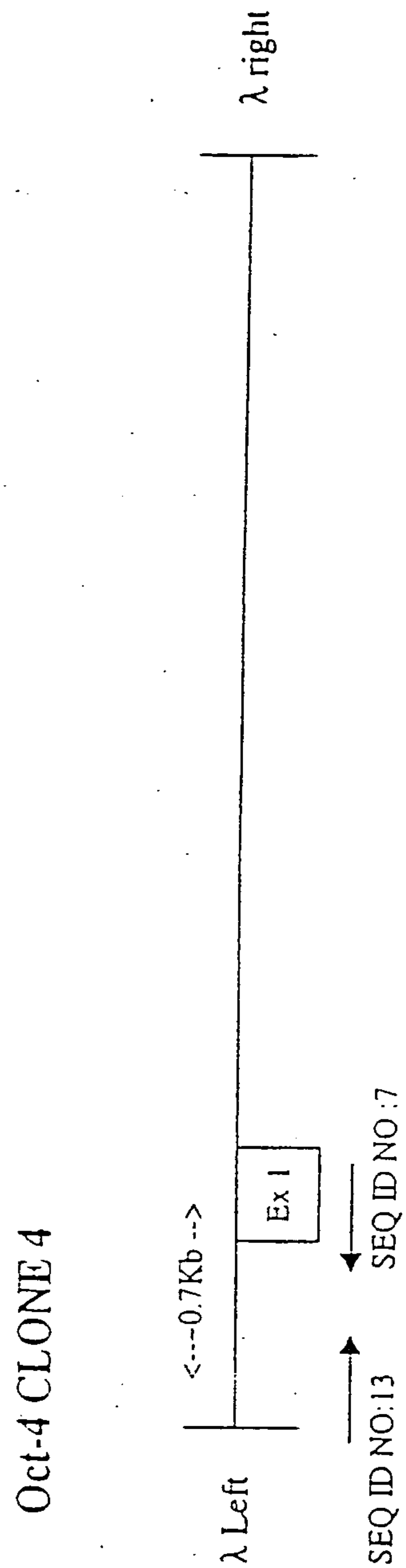

Oct-4 CLONE 3
λ right
λ Left
~7Kb
~4.2Kb
Ex 1
Ex 2
Ex 3
Ex 4
SEQ ID NO:7
SEQ ID NO:7
SEQ ID NO:7
SEQ ID NO:8
SEQ ID NO:15
SEQ ID NO:16
SEQ ID NO:13
SEQ ID NO:13

FIG. 5B



# FIG. 6A

```
                 10          20          30          40          50
         *        *     *     *     *     *     *     *     *     *
AGCCATGCTG GGTTGATCCT CCCACCTGGC TGTGCTTCCA AGGGCCTCCT

                 60          70          80          90         100
         *        *     *     *     *     *     *     *     *     *
GTTGGGTCAG GGATCTGGCG GGGGGCTGCT GGACCCAGAG GTGGGGAGGC

                110         120         130         140         150
         *        *     *     *     *     *     *     *     *     *
TTCTCTCATG CCCCCCGCCC TAGGACTTCT GCGGAGGGAT GGCCTACTGT

                160         170         180         190         200
         *        *     *     *     *     *     *     *     *     *
GCACCTCAGG TCAGAGAGGG GCTGGTGCCC CAAGGCGGCC TGGAGACCCC
```

FIG. 6B

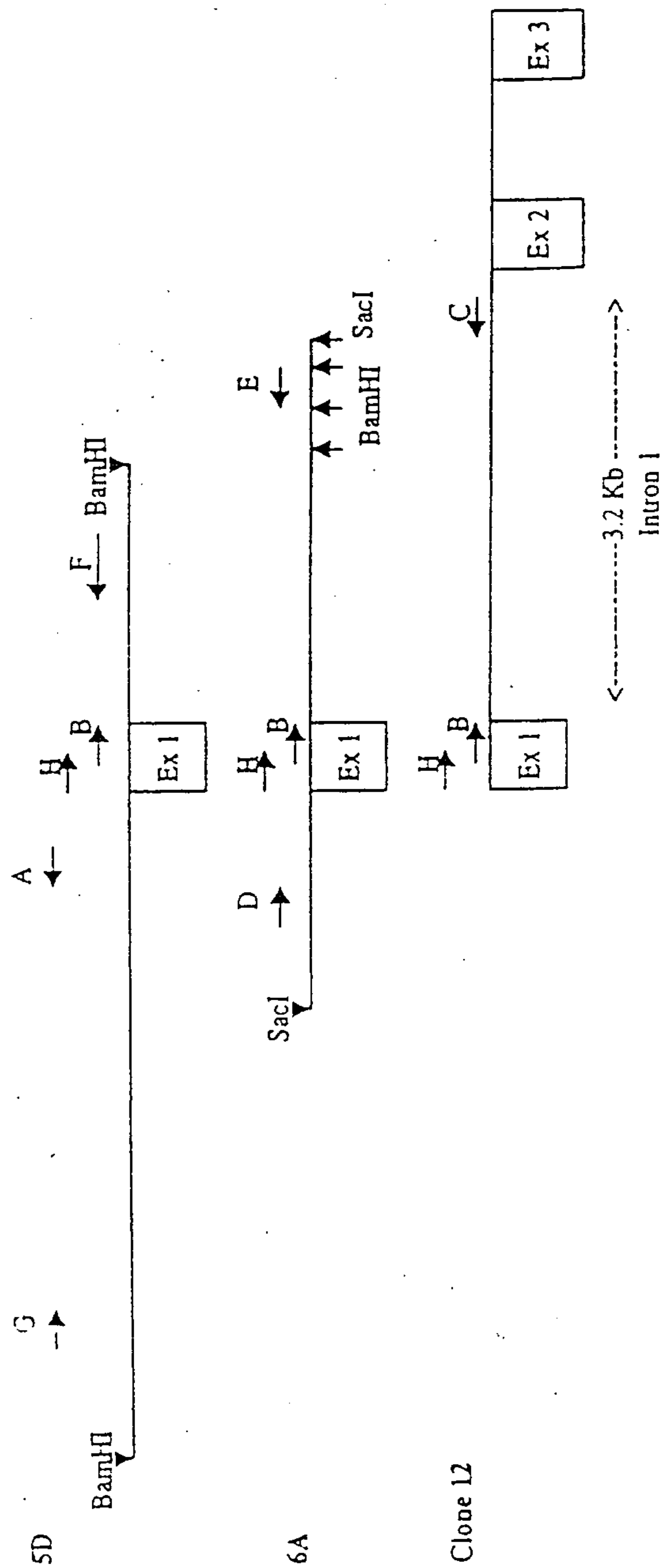
```
        210        220        230        240        250
    *    *     *    *     *    *     *    *     *    *
TCAGCCCTAG GGCCAGGCAG GAGTCGGGGT GGGGAGCAAC TCCGAGGGGG

        260        270        280        290        300
    *    *     *    *     *    *     *    *     *    *
CCTCCCTGGA GCCCTATGCC ACCCCCGTTG GCACTGCACA GCTGGACAAG

        310        320        330        340        350
    *    *     *    *     *    *     *    *     *    *
GAGAAACTAG AGCCGAATCC TGAGAAGTCC CAGGACATCA AAACGCTTCA

        360        370
    *    *     *    *     *
GAAAGACCTT CAACAATTTG CCAAGCTT
```

FIG. 7


5D
6A
Clone 12
BamHI
G
A
H
B
Ex 1
F
BamHI
SacI
D
E
SacI
C
Ex 2
Ex 3
3.2 Kb
Intron 1

# FIG. 8A

```
GGATCCCTAG CCTGGAAACC TCCATAAGCC GTGGGTACTG CCCTAAGAAG AAAAAAAAAG    60
TGGTTTGCTA CCCTGGTCTA GAGCAAGCCT CCATTTTCCC CAGGAGTCAT TTCAGCTGGT   120
TTTCCCTACC AAACAGCAAG GGGATGGCCG GGCTGACAGC AGCAAAGTCA CTGTCACCTC   180
TTTGCAGCCT TGCCAGGCCA GCTGCATCTG GCAGGGAGCG GCAGCTCTCA CCTGCCCTCC   240
CTGGGTCATG CTCTACAGCC AGATACTTGG CATTTGTCTT TGTGTAGGGC CTCAATATTG   300
TACTCTAATA AGGGTACATG TGGGAGTTCC CTGGTGGCTC AGCAGTTGAG GATCTGGCAT   360
TGTCACTGCT CTGGCATGGA TCTCTGCTCT GGCGCAGGTT CAATTCCTGG CCTGGGAACT   420
TCTGTACGCC GCAGGCGTGA CTGGAAAAAT ACAGGTGGGG TGGGGTGAGG AGTGTATGTG   480
GAGAGTCTGC AAACCCAGGC CTAAATTGGT TTGGGGGACT TGAAGTTTTT AGTGACTCCC   540
TACCCAAAAG AGTGGAGAAG CCAGGTCTGA TGACTTAACC CCACTTGCAG TCTGCTCTGG   600
GCCTGCAGAG ACCTGGCCTC TGCAGAAGTG AAGCTGCCTA CACTTCAGGC CTAACAGGAG   660
GGTGGGGAGG AGAGGGGAAT AGGCTCAGCC CTGCCATGCC AAGCACCCCC AGGCTGACTA   720
GGACTCCAGA CAAATTTAGC TTGTCCTTAA GGTTCTGGGT CAGACCCCAG GCAAGCACAG   780
AACTGATCTG GCTCAGATGT CTGGCTACAA GCTATCCAGG AACCCAGGCA TCCAGCCCTC   840
CCCAGCCCTC CCCAGGCTTT CCCTCTGGAA TAGGAAGGAC ACTTGCTTAA ACCAGAAACA   900
TACCATCTAG AGCAGCTATT TATGGTGATC TAAAAAACAC AGGGTGCTAT TTAGTCGGGG   960
GTGGGTGGGA AGGGAGAAGG TGTTTAGGGT CCGCGGGAAA GTCAGGGGCA CAGGGGCTCT  1020
CTGGACCACA TGGGGAGAGG GGTTTCTGGG AGGCCAGAGG GCGAGGAGCC AAGGAGCTCA  1080
```

FIG. 8B

```
GCAGTAGATT CCCTAGGCCC GCCCCTCCCC CTCCTCAGGG AGGCCGTCTT CTTGGCAGAC 1140
AGCAGATAGA TGCATGACAA AGGGGCCATG ATGGCTCTGT CCTGGGGGTT GGGGAGATGG 1200
CTAGGGAGGG GCCCCTCCTG GTCTGAAGCA CATCTTTCCA CCCCCACCAG GCCCCTTAAT 1260
CTATCTGCTT TTGGGGCAGT TAGTAGTTTA GAGTTGAAAT AGCTCCAGCC CTGCTGCCCT 1320
ATAAATCTTT CAACAGACCT ATGGGAAGTA TTGAAATGCA TGCACGCAAT TAGTCACCCC 1380
AAACGCACAG GCCGATGGGC ACTGGAAGAG ATTCAGAGGA GAAAAAGCAA AACAAAACAA 1440
CACGGACACA CAAAAACCCA ACAGACTCAA AGGACTCCTG GTGGAGCTAA CTGGTCACAG 1500
TCTGGAGGAT GCCAGCCCCT CAAGACAGAT GCCGAGCCAC TGACCCTAGC AAACAACCTC 1560
AGACCCAGCC AAGATGAAGA GGTGTCTAGG TCCGCAGAGG TCTGTGTCCC AGTCTCAGGA 1620
GTCTGGCCTC CAAACTGTAG GAAGCTCTGA TCCATGGCTT CTCTGGAGAG CCCCCCTCAC 1680
TCAGGTTCAC CTGGGGCCTT CGTTTAGGGC AAGTTGGGGG AGCAGACAGA CAAACATCAT 1740
CCCCAGCAGA CAGCCAGTCT GAAAGCTATT CTCTTGCAAA CAGAATCAAG CACTAGGCCA 1800
GCAGCCTGAG CCTCAGGACA GACCCAGAAA AATAGACCCT GTGGAGAGC TTAGGGCAGG 1860
ATTCCTGCAC CCCCTCCCCA ATCGCAGTTC ACCCCCTTCT GCATCTTTTC GCTAGCCCCC 1920
CAAACAAAGG CCTGGACGCC TCAGTCCTCT AGAGGGGGGA CAGGATACCT AGGTCCCAGT 1980
GGGGGGCCCT GTCTGAGGCT CAGTCTTTGA GGGGATGGGG GTGTTGTTGC TGGAGCTCTT 2040
TTAGCTGCTC TGAAGGGGAT TCTGTGTGAG GGGATTGGGG CTGGGGGGTT GGGGGGCAGG 2100
AAGCTGTCCC CAGGGGAGCC ATCCAGGCCC ATTCAAGGGT TGAGCACTTG TTTAGGGTTA 2160
```

FIG. 8C

```
GAGCTGCCCC CTCTGGGGAC CAGGATTGTC CAGCCAAGGC CATTGTCCGG CCCCCTTCCC  2220
CCAGTCCCTC CCAAGCCTCT TTGAACCTGA AGTCAGATAT TTTTCTCTCC CCCCCCCTCC  2280
CTCCTTGGCT TTCCCCACCC AGGGCCTAGG GGTGGAGGCC CAGATTGGGA GGTGGGGGAG  2340
GGAGAACAGT CAACTATGGG GCTAGATATT TGGGTCCCTG AAGGGGGGCT GGGGGACAAG  2400
GAACCTGATG TGCGCGGGGA CCCACAGCGG GGGACCTGCG AGCGGGTGTC CGATTGATTC  2460
CTCTGCCTGC ACAAAGATTG GGAGACTCAG GCCCAGTCCA TCGAGCTTGA TCCCTGGAAG  2520
GGAAAATGGG GGTTCCATCC CTGGGTCTGG TGGAAGGGAG GCCCCGGAAC CCGGAAAACT  2580
GTACGGAATG GAAGCCCGTG TGGCAGTCTG CCCCTGGTGA GGGGTGGAAT CTAATAGGCT  2640
GGGCGGATGG TTGCTGGGCA TCGCAGCTTT GGGGTGCCGG AATCTGGCCA GTAATCTAGT  2700
TGGGAATGCC TAGGTTCCCG GACTGGGGGT GAGGGCAGAG AGCAGGAATT GAGGAGTAGC  2760
TCCGGCAGGA CTTAGCACAG ACACCAGACC TGTGTGAGGA CCTGAGAGGG TCGCTGGGGT  2820
CCCTTGAGGA GACAGTGCCA GGGTCTTCGA AGAGGGGTCC AACACCTGGC TCCCCGACAG  2880
CCCCAATGTG CACAGAGCAG TGGAGAGGGC CGGGCGGCCG GTTGGGAGTT GGAGGTGAAG  2940
GCCGCATGGG GGACCTGCAC CAAGGGCCTG GGGACCGCAG AGGCGCCCGG GCGGACCTCT  3000
CCGACTTTCG CCCTCCAGAC ACCACCGCCA CCAGCCAGCA AACACCCTCC GCCTCAGTTT  3060
CTCCCACCCC CACCGACCCC TCCCCCACCC ATCCAGGGGG CGGGGCCAGA GGTCAAGGCT  3120
AGTGGGTGGG ATTGGGGAGG GAGAGAGGTG TCGAGCAGTC CCCTTGGAGA GCCCTGGTTT  3180
TACTGGGCCC CCGGCTTGGG GCGCCTTCCT TCCCC                             3215
```

# FIG. 9A

The Box indicated as A is the SP1/HRE domain

```
pig oct4 2701-3215   ---------- ---------- ---------- ---------- ----TGGGAA    6
huOCT4 1-499         ---------- ---------- ---------- ---------- ----------
mOCT4 1401-1950      TGCAATGGCT GTCTTGTCCT GGCCTTGGAC ATGGGCTGAA ATACTGGGTT   50

pig oct4 2701-3215   TGCCTAGGTT CCCGGACTGG GGGTGAGGGC AGAGAGC-AG GAAT--TGAG   53
huOCT4 1-499         ------GAAT TCAAGACC-A GCCTG-GGTA ACATAGCAAG GCCCCATCTC   42
mOCT4 1401-1950      CACCCATA-T CTAGGACTCT AGACG-GGTG GGTAAGCAA- GAAC--TGAG   95

pig oct4 2701-3215   GAGTAGCTCC GGCAGGACT- -TA--GCACA G-ACACCAG ACCTGT-GT-   95
huOCT4 1-499         TACTAAAAAT AAAAAAACTA ACAGGGCACA -GTGGTCCAA GCCTGTAGTC   91
mOCT4 1401-1950      GAGTGGCCCC AGAAA---TA AT-TGGCACA GAACATTCA A-TGGATGTT  140

pig oct4 2701-3215   -GAG-GACCT GAGAGG--GT CGCTG--GGG TCCCTTGAGG AGACAGT-GC  138
huOCT4 1-499         CCAGCCACTT AGGAGGCTGG AGCGGAAGGA TTGCTT-TGG -CCCAGTAGA  139
mOCT4 1401-1950      TTAGGCTCTC CAGAGGATG- -GCTGAGTGG GC-TGTAGG A--CAG--GC  183
```

FIG. 9B

```
pig oct4 2701-3215   C-AGGGTCTT CGAAGAGGGG TCCAACA--C CTGGCTCCCC GACAGCCCCA   185
huOCT4 1-499         T-CGGGGCTA CATTGAG--- -CCATCA--T TGTACTCCAC TGCA-CTCCA   181
mOCT4 1401-1950      CGAGGGGCTG C----AGTG- -CCAGCAGGC TTTGTGGTGC GATGGGGC--   225

pig oct4 2701-3215   ATGTGCACAG AGCAGTGGAG AGG-GCCGGG CGGCCGGTTG GGAGTTGGAG   234
huOCT4 1-499         GTCTGGGCAA CAAAGT-GAG A---CCCTGT C--TTAAAAA ATAAAAATAA   225
mOCT4 1401-1950      ATCCGAGCAA CTGGTTTGTG AGGTGTCCGG TGACCCA-AG GCAGGGGTGA   274

pig oct4 2701-3215   GTGAAGGCCG CATGGGGGAC CTGCACCAAG GGCCTGGGGA CCGCAGAGGC   284
huOCT4 1-499         AAAAAGTTTC TGTGGGGGAC CTGCACTGAG GTCCTGG--- ----AGGGGC   268
mOCT4 1401-1950      GAGCACCTTG AAGGTTGGAA ATG-AAGGCC TTCCTGGGGT CC-------C   316

pig oct4 2701-3215   GCCGGGCGGG ACCTCTCCGA CTTTCGCCTT C--CAGACAC CACCGCCACC   332
huOCT4 1-499         GCC-AGTTGT GTCTC-CCGG TTTTCCCCTT CCACAGACAC CATTGCCACC   316
mOCT4 1401-1950      GTCC-TAAGG GT-TGTCCTG -TCCAGACCT CCCCA---AC CTCCGTCTGG   360

pig oct4 2701-3215   AGC---CA-G CAAACAC-CC TCCGCCTCAG TTTCTCCCAC CCCCACCGAC   377
huOCT4 1-499         ACCAT-TACG CAAACATGCC TTCGCCTCAG TTTCTCCC-- -CCCACCTCC   362
mOCT4 1401-1950      AAGACACAGG CAGATA-GCG CTCGCCTCAG TTTCTCCACC CCC-ACAGCT   408
```

FIG. 9C

```
                                                  A
pig oct4 2701-3215   CC-CTCCTCC ACCCATCCAG GGGGCGGGGC CAGAGGTCAA GGCTAGTGGG   426
huOCT4 1-499         CT-CTCCTCC ACCCATCCAG GGGGCGGGGC CAGAGGTCAA GGCTAGTGGG   411
mOCT4 1401-1950      CTGCTCCTCC ACCCATCCAG GGGGCGGGGC CAGAGGTCAA GGCTAGAGGG   458

pig oct4 2701-3215   TGGGATTGGG GAGGGAGAGA GGTGTCGAGC AGTCCCTTTG GAGAGCCTTG   476
huOCT4 1-499         TGGGACTGGG GAGGGAGAGA GGGGTTGAGT AGTCCCTTTG CA-AGCCTTC   460
mOCT4 1401-1950      TGGGATTGGG GAGGGAGAG- ---GTGAAAC CGTCCC-TTG GTGAGCCTTC   503

pig oct4 2701-3215   GTTTTACTGG GCCCCCGGCT TGGGGTGCCT TCCTTCCCC- --------     515
huOCT4 1-499         ATTTCACCAG GCCCCCGGCT TGGGGTGCCT TCCTTCCCC- --------     499
mOCT4 1401-1950      TTTCCACCAG GCCCCCGGCT CGGGGTGCCC ACCTTCCCCA TGGCTGG      550
```

# FIG. 10A

```
Pig Oct-4 promoter   TCACAGTCTG GAGGATGCCA GCCCCTCAAG ACAGATGCCG AGCCACTGAC   1544
Mus S58422S1         ---------- ---------- ---------- ---------G AGCCACTGAC    440

Pig Oct-4 promoter   CCTAGCTAAC AACCTCAGAC CCAGCCAAGA TGAAGAGGTG TCTAGGTCCG   1594
Mus S58422S1         CCTAGCTAAC A-GCTCAG-- ---GC---G- -G--GCTGGG TCTAGG--CT    475

Pig Oct-4 promoter   CAGAGGTCTG TGTCCCGGTC TCAGGAGTCT GGCCTCCAAA CTGTGGGAAG   1644
Mus S58422S1         CAGAACTCTG T---CCTGGC T----A---T G-----TATA CTGTGGGGTG    510

Pig Oct-4 promoter   CTCTGATCCA TGGCTTCTCT GGAGAGCCCC CCTCATTCAG GTTCGCCTGG   1694
Mus S58422S1         CTCTG----- -GGCTT-TTT ---GAGGCTG TGTGATTCA- ---C-CCTGG    545

Pig Oct-4 promoter   GGCCTTCGTT TAGGGCAAGT TGGGGGAGCA GACAGACAAA CATCATCCCC   1744
Mus S58422S1         GGCCTTCGTT CAGGGCATGG TGTAGGAGCA GACAGACAAA CACCATCCCT    595

Pig Oct-4 promoter   AGCAGACAGC CAGTCTGAAA GCTATTCTCT TGCAAACAGA ATCAAGCACT   1794
Mus S58422S1         TGCAGACAGG CACTCTGAGG GCTATTCTCT TGCAAAGATA ACTAAGCACC    645
```

# FIG. 10B

```
Pig Oct-4 promoter  AGGCCAGCAG CCTGAGCCTC AGGACAGGCC CAGAAAAATA GACCCTGTGG  1844
Mus S58422S1        AGGCCAGTAA TGGGATCCTC AGACTGGGCC CAGAAAACCA ----CTGTTG   691

Pig Oct-4 promoter  GAGAGCTTAG GGCAGGTTTC CTGCACCCCC TCCCCAATCG CAGTTCACCC  1894
Mus S58422S1        GAAAGTTCAG GGTAGGCTCT CTGCACCCCC TCC----TCC TAATCCCGTC   737

Pig Oct-4 promoter  CCTTCTGCAT CTTTTCGCTA GCCCCCCAAA CAAGGGCCTG GAGGCCTCAG  1944
Mus S58422S1        TCTTTAGTCT CTTTCCGCCA GCAC----AG GAATGG--GG GAGGGGTGGG   781

Pig Oct-4 promoter  TCCTCTAGAG GGGGCACGGG A-TTCCT--A GG--TCCCAG T-GGGGGCCC  1988
Mus S58422S1        I-GACGAGGA TGAACACTGG AGTCCCTGGA GGAAGGGAAG CAGGGTATCT   830

Pig Oct-4 promoter  CTGTCTGAGG CTCGGTCTTT GAGGGGATGG GGGTGTTGTT GCTGGAGCTC  2038
Mus S58422S1        CCATCTGAGG CTCTGTCTTT GAGGTGA--- -GG------- --TGGAG---   864

Pig Oct-4 promoter  TTTTAGCTGC TCTGAAGGGG ATTCTGTGTG AGGGGATTGG GGCT---G-G  2084
Mus S58422S1        ----AGCTG- -GGGAAG--- -TCTTGTGTG AGGGGATTGG GGCTCAGG-G   904

Pig Oct-4 promoter  GGGGTTGGGG GGCAGGAAGC TGTCCCCAGG GGAGCCATCC GGGCCCATTC  2134
Mus S58422S1        GGGGTTGGGG TGCAGGAAGT TGTCCCCAGG GGAGCCATCC TGGCCCATTC   954
```

# FIG. 10C

```
Pig Oct-4 promoter   AAGGGTTGAG CACTTGTTTA GGGTTAGAGC TGCCCCATCT GGGGACCAGG   2184
Mus S58422S1         AAGGGTTGAG TACTTGTTTA GGGTTAGAGC TGCCCC-TCT GGGGACCAGG   1003

Pig Oct-4 promoter   ATTGTCCAGC CAAGGCCATT GTCCGGCCCC CTTCCCCCAG TCCCTCCCAA   2234
Mus S58422S1         ATTGTCCAGC CAAGGCCATT GTCCTGCCCC CTTCCCCCAG TCCCTCCCAG   1053

Pig Oct-4 promoter   GCCTCTTTGA ACCTGAAGTC AGATATTT-T TCTCTCCCCC CTCCTCCC-T   2282
Mus S58422S1         GCCTCTTTGA ACCTGAAGTC AGATATTTTT TCTCTCTACC CCCCTCCCAC   1103

Pig Oct-4 promoter   CC-TTGGCTT TCTCCACCCA GGGCTAGGG GTGGAGGCCC AGATTGGGAG   2331
Mus S58422S1         CCGTTGGCTT TCTCCACCCA GGACTAGG- CTGGAAGCTG GGATGAGGAG   1152

Pig Oct-4 promoter   GTGGGGGAGG GAGAACAGTC AACTATGGGG CTAGATATTT GGGTCCCTGA   2381
Mus S58422S1         GTGGGGGAGG GAGAAC---- ---------- ---------- ----------   1168
```

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/21289

**A. CLASSIFICATION OF SUBJECT MATTER**
IPC 6 C12N15/00 A01K67/027 C12N5/06 C12N5/10 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 A01K C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | WO 94 24274 A (UNIV EDINBURGH ;SMITH AUSTIN GERARD (GB); MOUNTFORD PETER SCOTT (G) 27 October 1994<br>cited in the application<br>see specially page 2 line 4<br>see the whole document<br>--- | 1 |
| Y | WO 95 20042 A (DALRYMPLE MICHAEL ALEXANDER ;PPL THERAPEUTICS SCOTLAND LTD (GB); S) 27 July 1995<br>see the whole document<br>--- | 1 |
| Y | WO 96 29395 A (STRINGER BRADLEY MICHAEL JOHN) 26 September 1996<br>see the whole document<br>--- | 1 |
| | -/-- | |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 24 February 1999 | 12/03/1999 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Chambonnet, F |

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | PIEDRAHITA J A ET AL: "ON THE ISOLATION OF EMBRYONIC STEM CELLS: COMPARATIVE BEHAVIOR OF MURINE, PORCINE AND OVINE EMBRYOS" THERIOGENOLOGY, vol. 34, no. 5, 1 November 1990, pages 879-901, XP000571055 | 1 |
| A | see the whole document | 4 |
| Y | EVANS M J ET AL: "DERIVATION AND PRELIMINARY CHARACTERIZATION OF PLURIPOTENT CELL LINES FROM PORCINE AND BOVINE BLASTOCYSTS" THERIOGENOLOGY, vol. 33, no. 1, January 1990, pages 125-128, XP002910077 | 1 |
| A | see the whole document | 4 |
| A | WO 97 25412 A (ANDERSON GARY B ;SHIM HOSUP (US); UNIV CALIFORNIA (US)) 17 July 1997 see the whole document | 1 |
| A | WHEELER M B: "DEVELOPMENT AND VALIDATION OF SWINE EMBRYONIC STEM CELLS: A REVIEW" REPRODUCTION, FERTILITY AND DEVELOPMENT, vol. 6, no. 5, 1994, pages 563-568, XP000607067 | 4 |
| A | YEOM, Y.I. ET AL.: "Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells" DEVELOPMENT, vol. 122, no. 3, 1996, pages 881-894, XP002094592 see the whole document | 1-4 |
| A | WIANNY, F. ET AL. : "A study of cyclin Ds expression during development in the mouse and pig: implication for embryonic stem cells derivation in pig" THERIOGENOLOGY, vol. 47, no. 1, 12 January 1997, XP002094593 see the whole document | 4 |

1

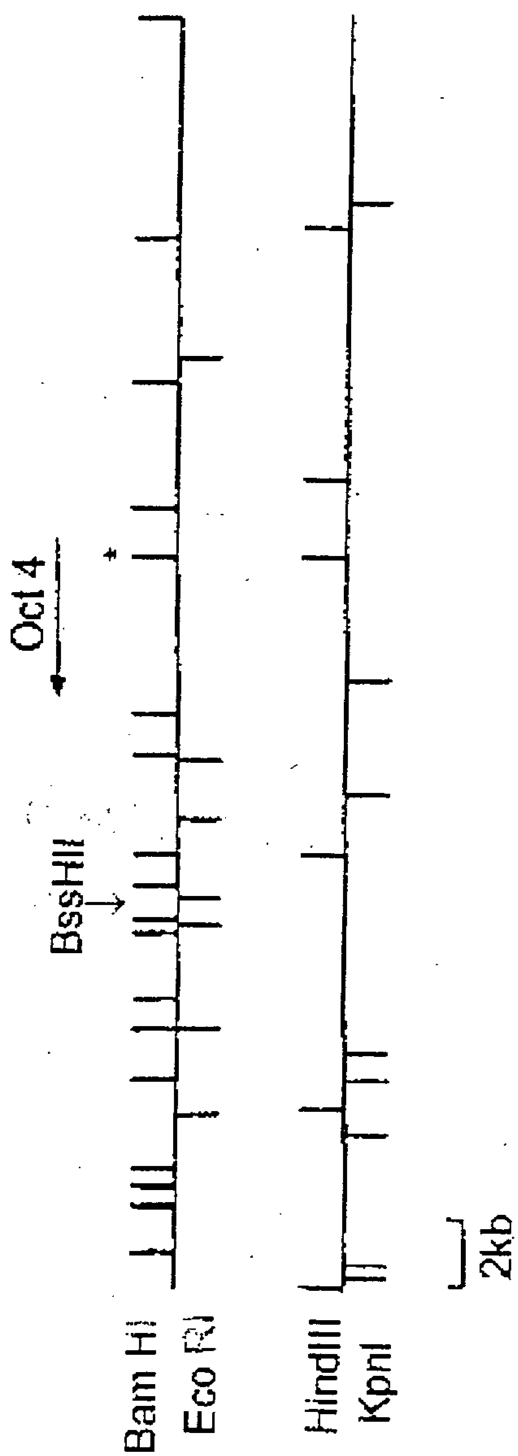
# INTERNATIONAL SEARCH REPORT

Information on patent family members

In ational Application No
PCT/US 98/21289

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 9424274 A | 27-10-1994 | AU 678233 B | 22-05-1997 |
| | | AU 6542694 A | 08-11-1994 |
| | | AU 678234 B | 22-05-1997 |
| | | AU 6542794 A | 08-11-1994 |
| | | CA 2161088 A | 27-10-1994 |
| | | CA 2161089 A | 27-10-1994 |
| | | EP 0695351 A | 07-02-1996 |
| | | EP 0695361 A | 07-02-1996 |
| | | WO 9424301 A | 27-10-1994 |
| | | JP 9500004 T | 07-01-1997 |
| | | JP 9500005 T | 07-01-1997 |
| | | NZ 265090 A | 24-03-1997 |
| | | NZ 265091 A | 27-07-1997 |
| | | SG 41951 A | 15-08-1997 |
| | | SG 48698 A | 18-05-1998 |
| | | ZA 9402719 A | 09-01-1995 |
| | | ZA 9402720 A | 30-03-1995 |
| WO 9520042 A | 27-07-1995 | AU 1461895 A | 08-08-1995 |
| | | EP 0741781 A | 13-11-1996 |
| | | JP 9508020 T | 19-08-1997 |
| WO 9629395 A | 26-09-1996 | AU 5116596 A | 08-10-1996 |
| | | CA 2214385 A | 26-09-1996 |
| | | CZ 9702959 A | 14-01-1998 |
| | | EP 0815206 A | 07-01-1998 |
| WO 9725412 A | 17-07-1997 | WO 9725413 A | 17-07-1997 |
| | | AU 1570697 A | 01-08-1997 |
| | | AU 4353696 A | 01-08-1997 |
| | | CA 2242666 A | 17-07-1997 |
| | | EP 0882127 A | 09-12-1998 |
| | | NZ 326996 A | 25-11-1998 |

FIG. 1A



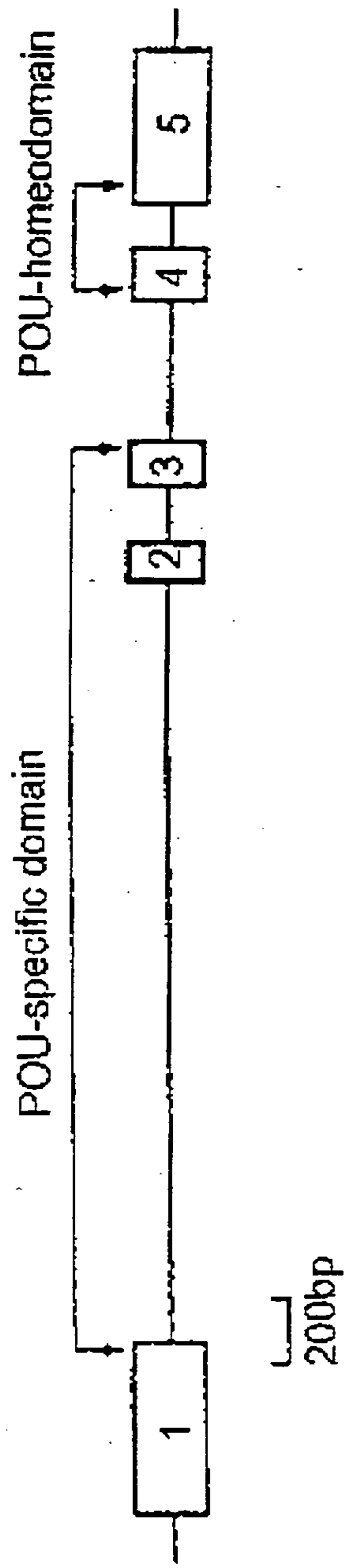



FIG. 1B

# FIG. 2A

```
                                              BamHI SmaI  PstI EcoRI
                                           AGTGGATCCCCCGGGCTGCAGGAATTCCGGG
      BamHI
      ATGGATCCTCGAACCTGGCTAAGCTTCCAAGGGCCTCCAGGTGGGCCTGGAATCGGACCA 60
  1   M  D  P  R  T  W  L  S  F  Q  G  P  P  G  G  P  G  I  G  P   -

      GGCTCAGAGGTATTGGGGATCTCCCCATGTCCGCCCGCATACGAGTTCTGCGGAGGGATG 120
 21   G  S  E  V  L  G  I  S  P  C  P  P  A  Y  E  F  C  G  G  M   -

      GCATACTGTGGACCTCAGGTTGGTCTGGGCCTAGTCCCCCAAGTTGGCGTGGAGACTTTG 180
 41   A  Y  C  G  P  Q  V  G  L  G  L  V  P  Q  V  G  V  E  T  L   -
                                              Clone 16
      CAGCCTGAGGGCCAGGCAGGAGCACGAGTGGAAAGCAACTCAGAGGGAACCTCCTCTGAG 240
 61   Q  P  E  G  Q  A  G  A  R  V  E  S  N  S  E  G  T  S  S  E   -

      CCCTGTGCCGACCGCCCCAATGCCGTGAAGTTGGAGAAGGTGGAACCAACTCCCGAGGAG 300
 81   P  C  A  D  R  P  N  A  V  K  L  E  K  V  E  P  T  P  E  E   -
                        PstI     Clone 1
      TCCCAGGACATGAAAGCCCTGCAGAAGGAGCTAGAACAGTTTGCCAAGCTGCTGAAGCAG 360
101   S  Q  D  M  K  A  L  Q  K  E  L  E  Q  F  A  K  L  L  K  Q   -
```

FIG. 2B

POU-SPECIFIC DOMAIN

```
    AAGAGGATCACCTTGGGGTACACCCAGGCCGACGTGGGGCTCACCCTGGGCGTTCTCTTT 420
121 K  R  I  T  L  G  Y  T  Q  A  D  V  G  L  T  L  G  V  L  F  -

    GGAAAGGTGTTCAGCCAGACCACCATCTGTCGCTTCGAGGCCTTGCAGCTCAGCCTTAAG 480
141 G  K  V  F  S  Q  T  T  I  C  R  F  E  A  L  Q  L  S  L  K  -

    AACATGTGTAAGCTGCGGCCCCTGCTGGAGAAGTGGGTGGAGGAAGCCGACAACAATGAG 540
161 N  H  C  K  L  R  P  L  L  E  K  W  V  E  E  A  D  N  N  E  -

    AACCTTCAGGAGATATGCAAATCGGAGACCCTGGTGCAGGCCCGGAAGAGAAAGCGAACT 600
181 N  L  Q  E  I  C  K  S  E  T  L  V  Q  A  R  K  R  K  R  T  -
```

POU HOMOEODOMAIN

```
    AGCATTGAGAACCGTGTGAGGTGGAGTCTGGAGACCATGTTTCTGAAGTGCCCGAAGCCC 660
201 S  I  E  N  R  V  R  W  S  L  E  T  M  F  L  K  C  P  K  P  -

    TCCCTACAGCAGATCACTCACATCGCCAATCAGCTTGGGCTAGAGAAGGATGTGGTTCGA 720
221 S  L  Q  Q  I  T  H  I  A  N  Q  L  G  L  E  K  D  V  V  R  -

    GTATGGTTCTGTAACCGGCGCCAGAAGGGCAAAAGATCAAGTATTGAGTATTCCCAACGA 780
241 V  W  F  C  N  R  R  Q  K  G  K  R  S  S  I  E  Y  S  Q  R  -
```

# FIG. 2C

```
    GAAGAGTATGAGGCTACAGGGACACCTTTCCCAGGGGGGGCTGTATCCTTTCCTCTGCCC 840
261 E  E  Y  E  A  T  G  T  P  F  P  G  G  A  V  S  F  P  L  P  -

    CCAGGTCCCCACTTTGGCACCCCAGGCTATGGAAGCCCCCACTTCACCACACTCTACTCA 900
281 P  G  P  H  F  G  T  P  G  Y  G  S  P  H  F  T  T  L  Y  S  -

    GTCCCTTTTCCTGAGGGCGAGGCCTTTCCCTCTGTTCCCGTCACTGCTCTGGGCTCTCCC 960
301 V  P  F  P  E  G  E  A  F  P  S  V  P  V  T  A  L  G  S  P  -

    ATGCATTCAAACTGAGGCACCAGCCCTCCCTGGGGATGCTGTGAGCCAAGGCAAGGGAGG 1020
321 M  H  S  N  *  324

    TAGACAAGAGAACCTGGAGCTTTGGGGTTAAATTCTTTTACTGAGGAGGGATTAAAAGCA 1080
    CAACAGGGGTGGGGGGTGGGATGGGGAAAGAAGCTCAGTGATGCTGTTGATCAGGAGCCT 1140
    GGCCTGTCTGTCACTCATCATTTTGTTCTTAAATAAAGACTGGGACACACAGTAAAAAAA 1200
    AAAAAAAAAACTCGAG 1217
              XhoI
```

FIG. 3A

```
        10         20         30         40         50
    *    *     *    *     *    *     *    *     *    *
GAATTCGGCT TCACCAGGCT TCGGACTTCG CCTTCTCGCC CCCGCCGGGC

        60         70         80         90        100
    *    *     *    *     *    *     *    *     *    *
GGTGGAGGCG ATGGGCCGGG AGGGCGGAGC CGGGCTGGGT TGATCCTCGG

       110        120        130        140        150
    *    *     *    *     *    *     *    *     *    *
ACCTGGCTGA GCTTCCAAGG GCCTCCCGGT GGGTCAGGGA TCGGGCCGGG

       160        170        180        190        200
    *    *     *    *     *    *     *    *     *    *
GGTTGGGCCG GGCGCCGAGG TGTGGGGGCT TCCCGCGTGC CCCCCGCCCT
```

# FIG. 3B

```
       210        220        230        240        250
   *     *    *     *    *     *    *     *    *     *
ATGACTTCTG CGGAGGGATG GCCTACTGCG CACCTCAGGT CGGAGTGGGG

       260        270        280        290        300
   *     *    *     *    *     *    *     *    *     *
CTGGTGCCCC AGGGCGGCCT GGAGACCCCT CAGCCCGAGG GCGAGGCGGG

       310        320        330        340
   *     *    *     *    *     *    *     *
GGCCGGGGTG GAGAGCAACT CCGAGGGGGC CTCCCCCGAG
```

# FIG. 4A

```
        10         20         30         40         50
    *    *     *    *     *    *     *    *     *    *
GGATCCTCGG ACCTGGCTGA GCTTCCAAGG GCCTCCCGGT GGGTCAGGGA

        60         70         80         90        100
    *    *     *    *     *    *     *    *     *    *
TCGGGCCGGG GGTTGGGCCG GGCGCCGAGG TGTGGGGGCT TCCCGCGTGC

       110        120        130        140        150
    *    *     *    *     *    *     *    *     *    *
CCCCCGCCCT ATGACTTCTG CGGAGGGATG GCCTACTGCG CACCTCAGGT

       160        170        180        190        200
    *    *     *    *     *    *     *    *     *    *
CGGAGTGGGG CTGGTGCCCC AGGGCGGCCT GGAGACCCCT CAGCCCGAGG
```

FIG. 4B

```
       210       220       230       240       250
    *    *    *    *    *    *    *    *    *    *
GCGAGGCGGG GGCCCGGGTG GAGAGCAACT CCGAGGGGGC CTCCCCCGAG

       260       270       280       290       300
    *    *    *    *    *    *    *    *    *    *
CCCTGTGCCG CCCCCGCTGG CGCCGCGAAG CTGGACAAGG AGAAGCTGGA

       310       320       330       340       350
    *    *    *    *    *    *    *    *    *    *
GCCGAACCCC GAAGAGGCGA GTGAGCTGCC GGGAGCTGGG GGAGGCGATC

       360       370       380       390       400
    *    *    *    *    *    *    *    *    *    *
GCGCTGGCCG GGGGCGCACG CAGGGGAGGT GGTCGCCTGC CGCCCCGGCA

    *
GGAGGGG
```

FIG. 5A

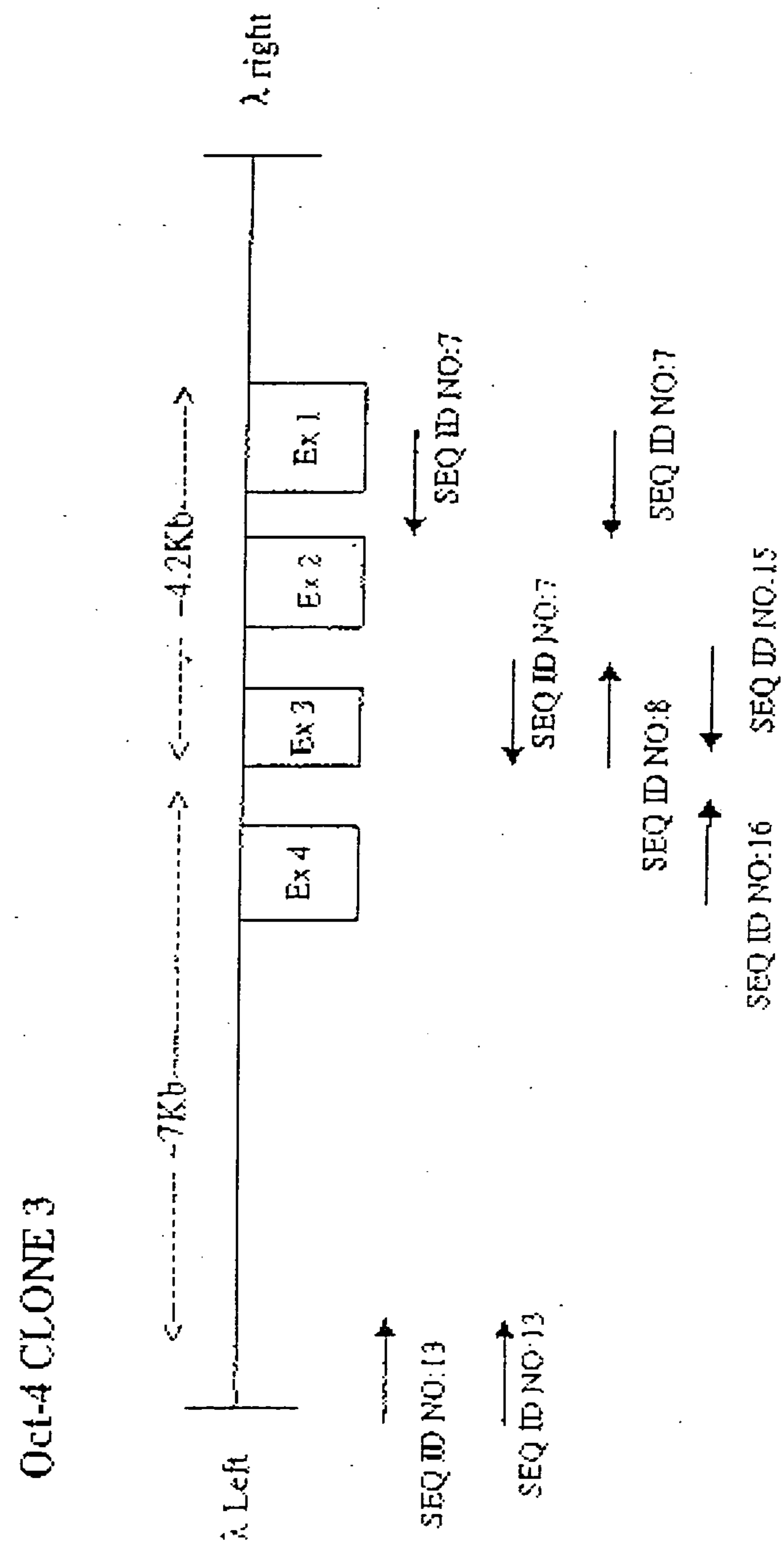
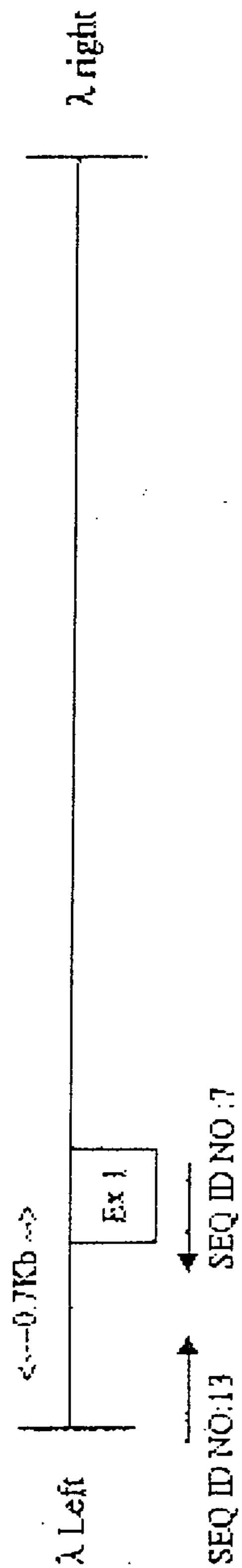

Oct-4 CLONE 3
λ Left
λ right
~7Kb
~4.2Kb
Ex 4
Ex 3
Ex 2
Ex 1
SEQ ID NO:13
SEQ ID NO:13
SEQ ID NO:7
SEQ ID NO:7
SEQ ID NO:8
SEQ ID NO:7
SEQ ID NO:16
SEQ ID NO:15

FIG. 5B

Oct-4 CLONE 4



# FIG. 6A

```
        10         20         30         40         50
         *          *          *          *          *
AGCCATGCTG GGTTGATCCT CCCACCTGGC TGTGCTTCCA AGGGCCTCCT

        60         70         80         90        100
         *          *          *          *          *
GTTGGGTCAG GGATCTGGCG GGGGGCTGCT GGACCCAGAG GTGGGGAGGC

       110        120        130        140        150
         *          *          *          *          *
TTCTCTCATG CCCCCCGCCC TAGGACTTCT GCGGAGGGAT GGCCTACTGT

       160        170        180        190        200
         *          *          *          *          *
GCACCTCAGG TCAGAGAGGG GCTGGTGCCC CAAGGCGGCC TGGAGACCCC
```

FIG. 6B

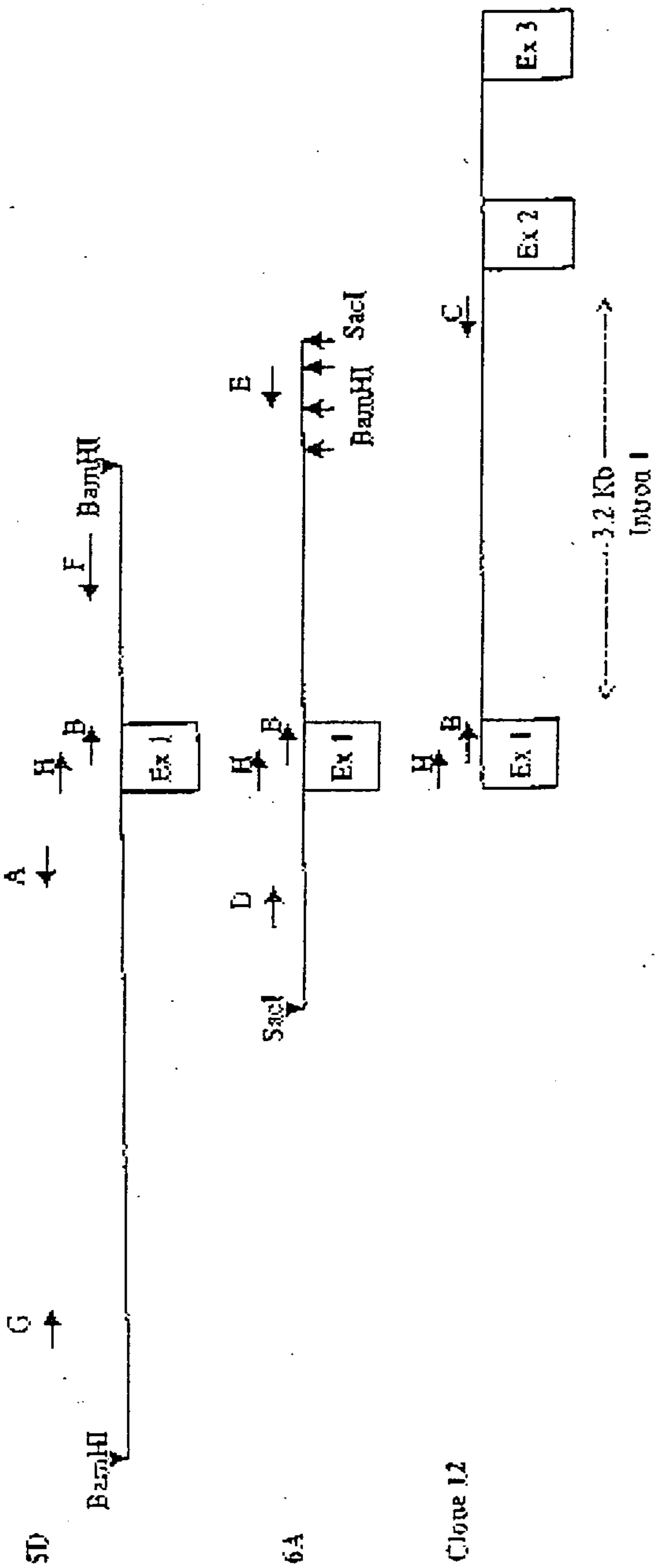
```
         210          220          230          240          250
    *    *     *     *     *     *     *     *     *     *
TCAGCCCTAG GGCCAGGCAG GAGTCGGGGT GGGGAGCAAC TCCGAGGGGG

         260          270          280          290          300
    *    *     *     *     *     *     *     *     *     *
CCTCCCTGGA GCCCTATGCC ACCCCCGTTG GCACTGCACA GCTGGACAAG

         310          320          330          340          350
    *    *     *     *     *     *     *     *     *     *
GAGAAACTAG AGCCGAATCC TGAGAAGTCC CAGGACATCA AAACGCTTCA

         360          370
    *    *     *     *     *
GAAAGACCTT CAACAATTTG CCAAGCTT
```

FIG. 7


5D
6A
Clone 12
BamHI
G
A
H
B
Ex 1
F
BamHI
SacI
D
E
BamHI
SacI
C
Ex 2
Ex 3
3.2 Kb
Intron 1

# FIG. 8A

| | | | | | | |
|---|---|---|---|---|---|---|
| GGATCCCTAG | CCTGGAAACC | TCCATAAGCC | GTGGGTACTG | CCCTAAGAAG | AAAAAAAAAG | 60 |
| TGGTTTGCTA | CCCTGGTCTA | GAGCAAGCCT | CCATTTTCCC | CAGGAGTCAT | TTCAGCTGGT | 120 |
| TTTCCCTACC | AAACAGCAAG | GGGATGGCCG | GGCTGACAGC | AGCAAAGTCA | CTGTCACCTC | 180 |
| TTTGCAGCCT | TGCCAGGCCA | GCTGCATCTG | GCAGGGAGCG | GCAGCTCTCA | CCTGCCCTCC | 240 |
| CTGGGTCATG | CTCTACAGCC | AGATACTTGG | CATTTGTCTT | TGTGTAGGGC | CTCAATATTG | 300 |
| TACTCTAATA | AGGGTACATG | TGGGAGTTCC | CTGGTGGCTC | AGCAGTTGAG | GATCTGGCAT | 360 |
| TGTCACTGCT | CTGGCATGGA | TCTCTGCTCT | GGCGCAGGTT | CAATTCCTGG | CCTGGGAACT | 420 |
| TCTGTACGCC | GCAGGCGTGA | CTGGAAAAAT | ACAGGTGGGG | TGGGGTGAGG | AGTGTATGTG | 480 |
| GAGAGTCTGC | AAACCCAGGC | CTAAATTGGT | TTGGGGGACT | TGAAGTTTTT | AGTGACTCCC | 540 |
| TACCCAAAAG | AGTGGAGAAG | CCAGGTCTGA | TGACTTAACC | CCACTTGCAG | TCTGCTCTGG | 600 |
| GCCTGCAGAG | ACCTGGCCTC | TGCAGAAGTG | AAGCTGCCTA | CACTTCAGGC | CTAACAGGAG | 660 |
| GGTGGGGAGG | AGAGGGGAAT | AGGCTCAGCC | CTGCCATGCC | AAGCACCCCC | AGGCTGACTA | 720 |
| GGACTCCAGA | CAAATTTAGC | TTGTCCTTAA | GGTTCTGGGT | CAGACCCCAG | GCAAGCACAG | 780 |
| AACTGATCTG | GCTCAGATGT | CTGGCTACAA | GCTATCCAGG | AACCCAGGCA | TCCAGCCCTC | 840 |
| CCCAGCCCTC | CCCAGGCTTT | CCCTCTGGAA | TAGGAAGGAC | ACTTGCTTAA | ACCAGAAACA | 900 |
| TACCATCTAG | AGCAGCTATT | TATGGTGATC | TAAAAAACAC | AGGGTGCTAT | TTAGTCGGGG | 960 |
| GTGGGTGGGA | AGGGAGAAGG | TGTTTAGGGT | CCGCGGGAAA | GTCAGGGGCA | CAGGGGCTCT | 1020 |
| CTGGACCACA | TGGGGAGAGG | GGTTTCTGGG | AGGCCAGAGG | GCGAGGAGCC | AAGGAGCTCA | 1080 |

# FIG. 8B

```
GCAGTAGATT CCCTAGGCCC GCCCCTCCCC CTCCTCAGGG AGGCCGTCTT CTTGGCAGAC 1140
AGCAGATAGA TGCATGACAA AGGGGCCATG ATGGCTCTGT CCTGGGGGTT GGGGAGATGG 1200
CTAGGGAGGG GCCCCTCCTG GTCTGAAGCA CATCTTTCCA CCCCCACCAG GCCCCTTAAT 1260
CTATCTGCTT TTGGGGCAGT TAGTAGTTTA GAGTTGAAAT AGCTCCAGCC CTGCTGCCCT 1320
ATAAATCTTT CAACAGACCT ATGGGAAGTA TTGAAATGCA TGCACGCAAT TAGTCACCCC 1380
AAACGCACAG GCCGATGGGC ACTGGAAGAG ATTCAGAGGA GAAAAAGCAA AACAAAACAA 1440
CACGGACACA CAAAAACCCA ACAGACTCAA AGGACTCCTG GTGGAGCTAA CTGGTCACAG 1500
TCTGGAGGAT GCCAGCCCCT CAAGACAGAT GCCGAGCCAC TGACCCTAGC AAACAACCTC 1560
AGACCCAGCC AAGATGAAGA GGTGTCTAGG TCCGCAGAGG TCTGTGTCCC AGTCTCAGGA 1620
GTCTGGCCTC CAAACTGTAG GAAGCTCTGA TCCATGGCTT CTCTGGAGAG CCCCCCTCAC 1680
TCAGGTTCAC CTGGGGCCTT CGTTTAGGGC AAGTTGGGGG AGCAGACAGA CAAACATCAT 1740
CCCCAGCAGA CAGCCAGTCT GAAAGCTATT CTCTTGCAAA CAGAATCAAG CACTAGGCCA 1800
GCAGCCTGAG CCTCAGGACA GACCCAGAAA AATAGACCCT GTGGGAGAGC TTAGGGCAGG 1860
ATTCCTGCAC CCCCTCCCCA ATCGCAGTTC ACCCCCTTCT GCATCTTTTC GCTAGCCCCC 1920
CAAACAAAGG CCTGGACGCC TCAGTCCTCT AGAGGGGGGA CAGGATACCT AGGTCCCAGT 1980
GGGGGGCCCT GTCTGAGGCT CAGTCTTTGA GGGGATGGGG GTGTTGTTGC TGGAGCTCTT 2040
TTAGCTGCTC TGAAGGGGAT TCTGTGTGAG GGGATTGGGG CTGGGGGGTT GGGGGGCAGG 2100
AAGCTGTCCC CAGGGGAGCC ATCCAGGCCC ATTCAAGGGT TGAGCACTTG TTTAGGGTTA 2160
```

# FIG. 8C

```
GAGCTGCCCC CTCTGGGGAC CAGGATTGTC CAGCCAAGGC CATTGTCCGG CCCCCTTCCC 2220
CCAGTCCCTC CCAAGCCTCT TTGAACCTGA AGTCAGATAT TTTTCTCTCC CCCCCCCTCC 2280
CTCCTTGGCT TTCCCCACCC AGGGCCTAGG GGTGGAGGCC CAGATTGGGA GGTGGGGGAG 2340
GGAGAACAGT CAACTATGGG GCTAGATATT TGGGTCCCTG AAGGGGGGCT GGGGGACAAG 2400
GAACCTGATG TGCGCGGGGA CCCACAGCGG GGGACCTGCG AGCGGGTGTC CGATTGATTC 2460
CTCTGCCTGC ACAAAGATTG GGAGACTCAG GCCCAGTCCA TCGAGCTTGA TCCCTGGAAG 2520
GGAAAATGGG GGTTCCATCC CTGGGTCTGG TGGAAGGGAG GCCCCGGAAC CCGGAAAACT 2580
GTACGGAATG GAAGCCCGTG TGGCAGTCTG CCCCTGGTGA GGGGTGGAAT CTAATAGGCT 2640
GGGCGGATGG TTGCTGGGCA TCGCAGCTTT GGGGTGCCGG AATCTGGCCA GTAATCTAGT 2700
TGGGAATGCC TAGGTTCCCG GACTGGGGGT GAGGGCAGAG AGCAGGAATT GAGGAGTAGC 2760
TCCGGCAGGA CTTAGCACAG ACACCAGACC TGTGTGAGGA CCTGAGAGGG TCGCTGGGGT 2820
CCCTTGAGGA GACAGTGCCA GGGTCTTCGA AGAGGGGTCC AACACCTGGC TCCCCGACAG 2880
CCCCAATGTG CACAGAGCAG TGGAGAGGGC CGGGCGGCCG GTTGGGAGTT GGAGGTGAAG 2940
GCCGCATGGG GGACCTGCAC CAAGGGCCTG GGGACCGCAG AGGCGCCCGG GCGGACCTCT 3000
CCGACTTTCG CCCTCCAGAC ACCACCGCCA CCAGCCAGCA AACACCCTCC GCCTCAGTTT 3060
CTCCCACCCC CACCGACCCC TCCCCCACCC ATCCAGGGGG CGGGGCCAGA GGTCAAGGCT 3120
AGTGGGTGGG ATTGGGGAGG GAGAGAGGTG TCGAGCAGTC CCCTTGGAGA GCCCTGGTTT 3180
TACTGGGCCC CCGGCTTGGG GCGCCTTCCT TCCCC                           3215
```

# FIG. 9A

The Box indicated as A is the SP1/HRE domain

```
pig oct4 2701-3215   ---------- ---------- ---------- ---------- ----TGGGAA     6
huOCT4 1-499         ---------- ---------- ---------- ---------- ----------
mOCT4 1401-1950      TGCAATGGCT GTCTTGTCCT GGCCTTGGAC ATGGCTGAA ATACTGGGTT     50

pig oct4 2701-3215   TGCCTAGGTT CCCGGACTGG GGGTG-GGGC AGAGAGC-AG GAAT--TGAG    53
huOCT4 1-499         ------GAAT TCAAGACC-A GCCTG-GGTA ACATAGC-AG GCCCCATCTC    42
mOCT4 1401-1950      CACCCATA-T CTAGGACTCT AGACG-GGTG GGTAAGCAA- GAAC--TGAG    95

pig oct4 2701-3215   GAGTAGCTCC GGCAGGACT- -TA--GCACA -G-ACACCAG ACCTGT-GT-    95
huOCT4 1-499         TAGTAAAAAT AAAAAAACTA ACAGGGCACA -GTGGTCCAA GCCTGTAGTC    91
mOCT4 1401-1950      GAGTGGCCCC AGAAA---TA AT-TGGCACA -GAACATTCA A-TGGATGTT   140

pig oct4 2701-3215   -GAG-GAGCT GAGGAGG--GT CGCTG--GGG TCCCTTGAGG AGACAGT-GC  138
huOCT4 1-499         CCAGCCAGTT AGGAGGCTGG AGCCGAAGGA TTGCTT-TGG -CCCAGTAGA   139
mOCT4 1401-1950      TAAGGCTGTC CAGAGGATG- -GCTGAGTGG GC-TGTAAGG A--CAG--G-   183
```

# FIG. 9B

```
pig oct4 2701-3215   C-AGGCTCTT CGAAGAGGGG TCCAGCA--C CTGGCTCCCC GACAGCCCCA   185
huOCT4 1-499         T-CGGGGCTA CATTGAG--- -CCAGCA--T TGTACTCCAC TGCA-CTCCA   181
mOCT4 1401-1950      CGAGGGGCTG C----AGTG- -CCAGCAGGC TTTGTGGTGC GATGGGGC--   225

pig oct4 2701-3215   ATGTGCACAG AGCAGTGGGG AGG-GCCGCG CCGCCGGTTG GGAGTTGGAG   234
huOCT4 1-499         GTTTGGGCAA CAAAGTTGGG A---CCCGGT C--TTAAAAA ATAAAAATAA   225
mOCT4 1401-1950      ATCTGAGCAA CTGGTTTGTG AGGTGTCCGG TGACCCA-AG GCAGGGGTGA   274

pig oct4 2701-3215   GTGAAGGCCG CATGGGGGAC CTGGACCAAG GGCCTGGGGA CCGCAGAGGC   284
huOCT4 1-499         AAAAAGTTTC TGTGGGGGAC CTGGACTGAG GTCCTGG--- -----AGGGC   268
mOCT4 1401-1950      GAGGACCTTG AAGGTTGGAA ATGGAGGCC TTCCTGGGGT CC-------C   316

pig oct4 2701-3215   GACCGGGCGG ACCTCTCCGA CTTTCGCCTT C--CAGACAC CACCGCCACC   332
huOCT4 1-499         GGC-AGTTGT GTCTC-CCGG TTTTCCCCTT CCACAGACAC CATTGCCACC   316
mOCT4 1401-1950      GGCC-TAAGG GT-TGTCCTG -TCCAGACTT CCCCA---AC CTCCGCTTGG   360

pig oct4 2701-3215   AGC---CAAG CAAAGAC-CC TCCGCCTCAG TTTCTCCCAC CCCCACCGAC   377
huOCT4 1-499         ACCAT-TAGG CAAACAATCC TTCGCCTCAG TTTCTCCC-- -CCCACCTCC   362
mOCT4 1401-1950      AAGACACAGG CAAATA-CCG CTCGCCTCAG TTTCTCCACC CCCCACAGCT   408
```

# FIG. 9C

```
                                                     A
pig oct4 2701-3215   CT-CTCCTCC ACCCATCCAG GGGGCGGGGC CAGAGGTCAA GGCTAGTGGG   426
huOCT4 1-499         CT-CTCCTCC ACCCATCCAG GGGGCGGGGC CAGAGGTCAA GGCTAGTGGG   411
mOCT4 1401-1950      CTGCTCCTCC ACCCACCCAG GGGGCGGGGC CAGAGGTCAA GGCTAGAGGG   458

pig oct4 2701-3215   TGGGAATGGG GAGGGAGACA GGTGTTGAGC AGTCCCTTTG GAGAGCCTTG   476
huOCT4 1-499         TGGGAATGGG GAGGGAGACA GGGGTTGAGT AGTCCCTTTG CA-AGCCCTC   460
mOCT4 1401-1950      TGGGAATGGG GAGGGAGAG- ---GTTAAAC CGTCCC-TTG GTCAGCCTTC   503

pig oct4 2701-3215   CTTTTACTGG GCCCCCGGCT TGGGGTGCCT TCCTTCCCC- --------     515
huOCT4 1-499         ATTTTACTAG GCCCCCGGCT TGGGGTGCCT TCCTTCCCC- --------     499
mOCT4 1401-1950      TTTCCACCAG GCCCCCGGCT GGGGGTGCCC ACCTTCCCCA TGGCTGG      550
```

# FIG. 10A

```
Pig Oct-4 promoter   TCACAGTCTG GAGGATGCCA GCCCCTCAAG ACAGATGCCG AGCCACTGAC   1544
Mus S58422S1         ---------- ---------- ---------- ---------G AGCCACTGAC    440

Pig Oct-4 promoter   CCTAGCAAAC AACCTCAGAC CCAGCCAACA TGAAGAGGTG TCTAGGTCTG   1594
Mus S58422S1         CCTAGCCAAC A-GCTCAG-- ---GC---G- -G--GCTGGG CCTAGG--CT    475

Pig Oct-4 promoter   CAGAGGTCTG TGTCCCAGCC TCAGGAGTCT GGCCTCCAAA CTGTGGCAAG   1644
Mus S58422S1         CAGAACTCTG T---CCTGGC T----A---T G------TAA CTGTGGGTGG    510

Pig Oct-4 promoter   CTCTGATCCA TGGCTTCTCT GGAGAGCTCC CCTCACTCAG GTTCCCCTGG   1694
Mus S58422S1         CTCTG----- -GGCTTCTTT ---GAGGCTG TGTGATTCA- ---C-CCTGG    545

Pig Oct-4 promoter   GGCCTTCGTT CAGGGCAAGT TGGGGGAGCA GACAGACAAA CAGCATCCCC   1744
Mus S58422S1         GGCCTTCGTT CAGGGCATGG TGTAGGAGCA GACAGACAAA CAGCATCCCT    595

Pig Oct-4 promoter   AGCAGACAGC CATTCTGAAA GCTATTCTCT TGCAAAGAGA ATCAAGCACT   1794
Mus S58422S1         TGCAGACAGG CACTCTGAGG GCTATTCTCT TGCAAAGATA ACTAAGCACC    645
```

# FIG. 10B

| | | |
|---|---|---|
| Pig Oct-4 promoter | AGGCCAGGAG CCTGAGCCTC AGGACAGGCC CAGAAAAATA GACCCTGTGG | 1844 |
| Mus S58422S1 | AGGCCAGGAA TGGGATCCTC AGACTGGGCC CAGAAAATCA ----CTTTTG | 691 |
| Pig Oct-4 promoter | GAGAGTTTAG GGGAGGGTTC CTGCACCCCC TCCCCAATCG CAGTTTCACC | 1894 |
| Mus S58422S1 | GATAGTCTAG GGTAGGCTCT CTGCACCCCC TCC----TCC TAATTCCGTC | 737 |
| Pig Oct-4 promoter | CGTTCTGCAT CTTTTCGCTA GCTCCCCAAA CAAAGGCCTG GAGGCCTCAG | 1944 |
| Mus S58422S1 | TGTTCAGTCT CTTTCCGCTA GCTC----AG CAATGG--GG GAGGGCTGGG | 781 |
| Pig Oct-4 promoter | TTCTCTAGAG GGGGGACTGG ATTCCCT--A GG--TCCCAG T-GGGGGGCC | 1988 |
| Mus S58422S1 | T-GACTAGGA TGAACACTGG AGTTCCTGGA GGAAGGGAAG CAGGGTATTT | 830 |
| Pig Oct-4 promoter | CTGTCTGAGG CTCAGTCTTT GAGGGGATGG GGGTGTTGTT GCTGGAGCTC | 2038 |
| Mus S58422S1 | CCATCTGAGG CTCTGTCTTT GAGGAGA--- -GG------- --TGGAG--- | 864 |
| Pig Oct-4 promoter | TTTTAGCTGC TCTGAAGGGG ATTCTGTGTG AGGGGATTGG GGCT---G-G | 2084 |
| Mus S58422S1 | ----AGCTG- -GGGAAG--- -TCTTGTGTG AGGGGATTGG GGCTCAGG-G | 904 |
| Pig Oct-4 promoter | GGGGTTGGGG GGCAGGAAGC TGTCCCCAGG GGAGCCATCC AGGCCCATTC | 2134 |
| Mus S58422S1 | GGGGTTGGGG GGCAGGAAGT TGTCCCCAGG GGAGCCATCC TGGCCCATTC | 954 |

# FIG. 10C

| | | |
|---|---|---|
| Pig Oct-4 promoter | AAGGGTTGAG TACTTGTTTA GGGTTAGAGC TGCCCCTTCT GGGGACCAGG | 2184 |
| Mus S58422S1 | AAGGGTTGAG TACTTGTTTA GGGTTAGAGC TGCCCC-TCT GGGGACCAGG | 1003 |
| Pig Oct-4 promoter | ATTGTCCAGC CAAGGCCATT GTCCTGCCCC CTTCCCCCAG TCCCTCCCAA | 2234 |
| Mus S58422S1 | ATTGTCCAGC CAAGGCCATT GTCCTGCCCC CTTCCCCCAG TCCCTCCCAG | 1053 |
| Pig Oct-4 promoter | GCCTCTTTGA ACCTGAAGTC AGATATTTTT TCTCTCCCCC CCCCTCCC-T | 2282 |
| Mus S58422S1 | GCCTCTTTGA ACCTGAAGTC AGATATTTTT TCTCTCTACC CACCTCCCAC | 1103 |
| Pig Oct-4 promoter | CC-TTGGTTT TCCCACCCCA GGGCCTAGGG GTGGAGGCCC AGATTGGGAG | 2331 |
| Mus S58422S1 | CCTTTGGTTT TCTCCACCCA GGAACTAGG- GTGGAGGCTG GGATGAGGAG | 1152 |
| Pig Oct-4 promoter | GTGGGGGAGG GAGAACAGTC AACTATGGGG CTAGATATTT GGGTCCCTGA | 2381 |
| Mus S58422S1 | GTGGGGGAGG GAGAAC---- ---------- ---------- ---------- | 1168 |